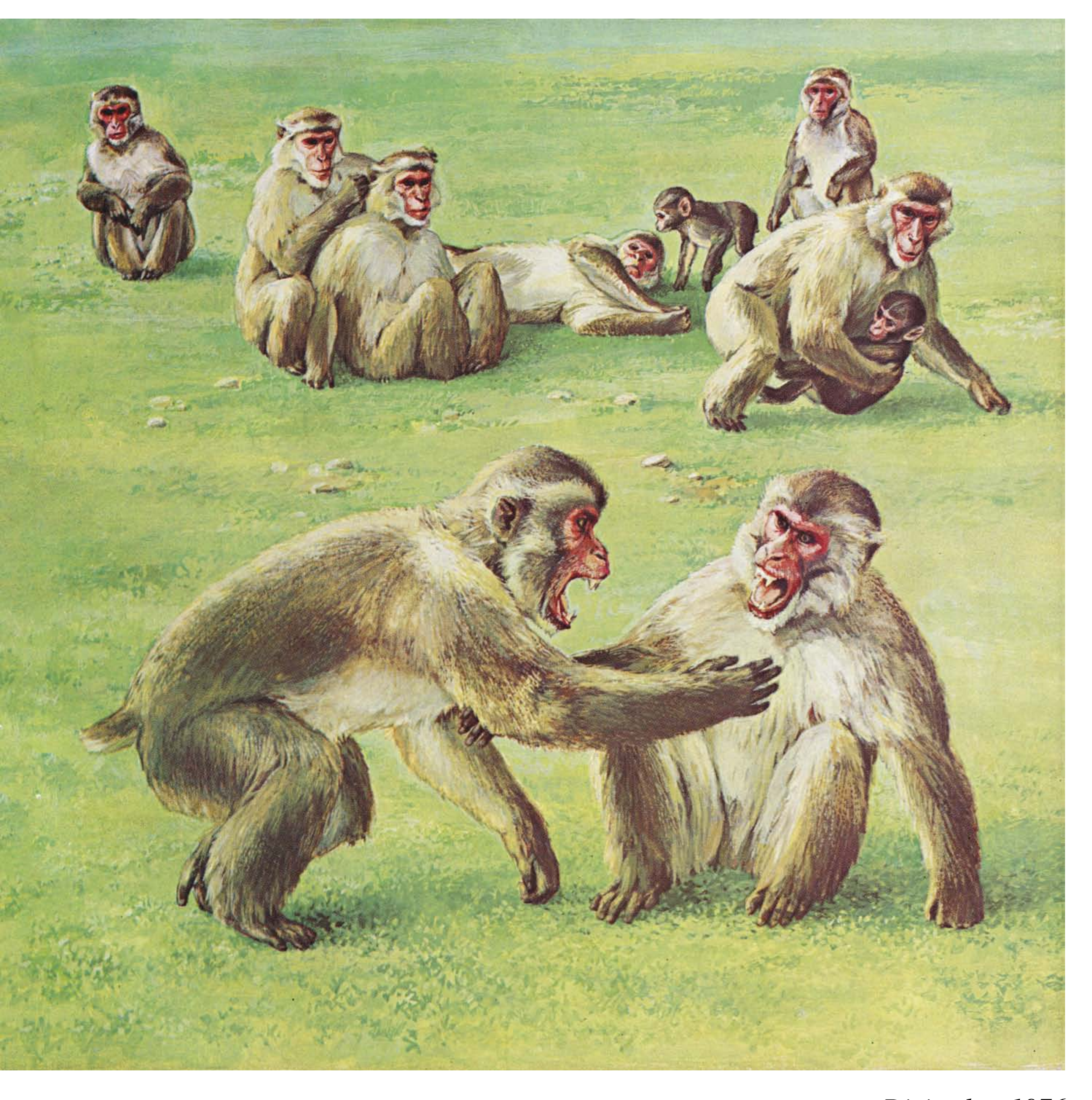
INVESTIGACION Y CIENCIA

Edición en español de **SCIENTIFIC AMERICAN**



ORGANIZACION SOCIAL DE LOS MACACOS JAPONESES

Diciembre 1976

140 Ptas.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 3
Diciembre 1976

INVESTIGACION Y CIENCIA

Edición en español de **SCIENTIFIC AMERICAN**

EDITA Prensa Científica, S.A.
Calabria, 235-239
Barcelona-15

Colaboradores de este número:

Asesoramiento y traducción:

Xavier Llimona: *La erradicación de la viruela;* José Millán Gómez: *La generación fotovoltaica de electricidad;* Ramón Fontarnau Griera: *Estudios sobre los ribosomas mediante dispersión de neutrones;* Ramón Canal Masgoret: *Erupciones de rayos gamma cósmicos;* Juan Santiago Muñoz: *Hologramas de luz blanca;* José Egozcúe: *Organización social de los macacos japoneses;* Manuel Puigcerver: *Tempestades de polvo;* José M.ª Vidal Llenas: *Taller y laboratorio;* Enrique Trillas y Teresa Riera: *Juegos matemáticos.*

Ciencia y sociedad:

Roser González; Arturo Caballero: Francisco Gil

Libros:

Violeta Demonte



LA PORTADA

La reproducción de la portada muestra nueve miembros del grupo de 230 macacos japoneses que viven en condiciones seminaturales en el Oregon Regional Primate Research Center. En primer plano se ve el principio de una lucha entre dos macacos machos adultos, hecho frecuente durante la fase reproductiva de otoño e invierno. A la derecha, una hembra adulta se lleva a su cría lejos del área. Detrás de ella y en la parte izquierda, están sentados dos machos pertenecientes al rango inferior. En medio del grupo un pequeño se aproxima a un joven que está dormitando y a una pareja consorte; el macho de la pareja, alarmado por los gritos de los monos en lucha, ha parado momentáneamente de espulgar a la hembra. El comportamiento de los macacos es de especial interés, al no venir determinado por factores biológicos y adquirir gran importancia los factores sociales y de desarrollo, dando lugar a una estructura social flexible capaz de adaptarse a múltiples situaciones y ambientes varios. G. Gray Eaton ha estado estudiando el comportamiento de estos monos durante los últimos seis años. Describe sus observaciones en la página 70.

Suscripciones:

Prensa Científica, S.A.
Departamento de suscripciones
Calabria, 235-239
Barcelona-15 (España)

Distribución para España:

Distribuciones de Enlace, S.A.
Ausias March, 49
Barcelona-10 (España)

Distribución para los restantes países:

Editorial Labor, S.A.
Calabria, 235-239
Barcelona-15 (España)

Condiciones de suscripción:

España:
Un año (12 números): 1.540 pesetas
Seis meses (6 números): 800 pesetas

Extranjero:
Un año (12 números); 30 U.S. $

Ejemplar atrasado ordinario:
160 pesetas

Ejemplar atrasado extraordinario:
225 pesetas.

PROCEDENCIA DE LAS ILUSTRACIONES

Diseño de la cubierta por Peter McGinn

Página	Fuente	Página	Fuente
4	World Health Organization	73	Andrew Christie
5-10	Lorelle M. Raboni	74-75	Sarah Landry
11	Fritz Goro World Health Organization	76-77	Andrew Christie
		78-79	Sarah Landry
		80	Andrew Christie
12	World Health Organization		
16-17	Ralph Morse	82-83	Bradshaw Photo, Big Spring, Tex.
18-25	Dan Todd	84	Alan D. Iselin
29	James A. Lake	85	Hamilton E. McRae III
30-38	Bunji Tagawa	86	Sherwood B. Idso Con Keyes, EPA-Documerica
45-53	Allen Beechel		
57-60	Fritz Goro	87	National Oceanic and Atmospheric Administration Theodore F. Tyler, U.S. Geological Survey
61-67	George V. Kelvin		
70	Kurt B. Modahl Jon Brenneis		
		91-98	Manuel Ballester
71	Andrew Christie		
72	Jon Brenneis	101-104	Andrew Christie

Dep. legal: B. 38.999-76
Fotocomposición: Fort, S.A.
Rosellón, 33 - Barcelona-15
Cayfosa. Santa Perpetua de Moguda.
Barcelona.
Printed in Spain - Impreso en España

Los autores

DONALD A. HENDERSON ("La erradicación de la viruela") es jefe médico, encargado de la erradicación de la viruela en la OMS. Graduado en el Oberlin College, recibió su M.D. en la Facultad de Medicina y Odontología de la Universidad de Rochester, en 1954 y su M.P.H. en la Escuela de Higiene y Salud Pública de la Johns Hopkins University, en 1960. Estuvo adscrito durante varios años al Epidemic Inteligence Service, del Communicable Disease Center del Department of Health, Education and Welfare; en calidad de jefe de la sección de Vigilancia, desde 1961 hasta 1966. En noviembre de 1965, cuando el presidente Johnson creó el programa de ayuda para la erradicación de la viruela a 18 países del Africa occidental y central, fue nombrado director.

BRUCE CHALMERS ("La generación fotovoltaica de electricidad") es profesor de Metalurgia en la Universidad de Harvard y consultor del programa de desarrollo de la célula solar de silicio de la Mobil Tyco Solar Energy Corporation. Estudió en la Universidad de Londres, graduándose en 1929 y doctorándose en 1932 en Física. Partiendo de su interés original por la investigación de la física de la deformación plástica, acometió el estudio del crecimiento de monocristales de metales.

DONALD M. ENGELMAN y PETER B. MOORE ("Estudios sobre los ribosomas mediante dispersión de neutrones") son ambos profesores agregados en el departamento de Biofísica y Bioquímica molecular de la Universidad de Yale. Engelman recibió su B.A. en el Reed College, en 1962, y su Ph.D. en Biofísica en Yale, en 1967. Pasó el siguiente año como postdoctorado en el laboratorio Walther Stoeckenius del Hospital Clínico de la Universidad de California en San Francisco; durante dos años fue becario del King's College, de la Universidad de Londres, en donde trabajó con M.H.F. Wilkins y John Randall. Regresó a Yale en 1970 como profesor ayudante. Moore recibió su B.S. summa cum laude en Yale, en 1961, y su Ph.D. en la Universidad de Harvard, en 1966, ambos en Biofísica. Realizó su tesis en el laboratorio James Watson. Pasó luego un año como postdoctorado en el Instituto de Biología molecular de la Universidad de Ginebra. En 1968 trabajó en el Medical Research Council Laboratory of Molecular Biology, de Cambridge, Inglaterra, con H. E. Huxley.

IAN B. STRONG y RAY W. KLEBESADEL ("Erupciones de rayos gamma cósmicos") forman parte del personal del laboratorio científico de Los Alamos (Estados Unidos). Strong nació en el estado de Nueva York, pero se crió en Escocia. Cursó estudios en la Universidad de Glasgow, obteniendo su B.Sc. en Física en 1953. Después de algunos años en los laboratorios Bell, obtuvo su Ph. D. en Física nuclear por la Universidad de Pennsylvania. Klebesadel se crió en una pequeña granja de Wisconsin. Tras cuatro años de servicio en las Fuerzas Aéreas, acudió a la Universidad de Wisconsin, donde recibió su B.A. en Ingeniería eléctrica en 1959. Se encargó de la supervisión de la instalación del acelerador tándem Van de Graaf de Wisconsin, con lo cual obtuvo un M.S. en Física.

EMMETT N. LEITH ("Hologramas de luz blanca") es investigador jefe en el Willow Run Laboratories de la Universidad de Michigan (Estados Unidos). Se licenció en 1949 en la Wayne State University, obteniendo su Master en Ciencias en 1952. Ese mismo año se incorporó al Willow Run Laboratories. En 1955, cuando él y sus colaboradores estaban trabajando en un sistema de imágenes de radar, descubrieron un nuevo modo de grabar imágenes de microondas sobre películas fotográficas, una técnica semejante en muchos aspectos a los primeros ensayos de Dennis Gabor en holografía, desconocidos entonces para ellos. Algunos años más tarde, Leith y su colega Juris Upatnieks demostraron por primera vez las posibilidades de la holografía para formar imágenes tridimensionales de alta calidad.

G. GRAY EATON ("Organización social de los macacos japoneses") trabaja como científico asociado del Oregon Regional Primate Research Center de Beaverton, Oregon. Nacido en Canadá, obtuvo su B.A. en Psicología en la Universidad de Victoria en 1964. A continuación, se trasladó a la Universidad de California, en Berkeley, iniciando una serie de estudios sobre lesiones cerebrales en primates. "Beach me aceptó como investigador, y mi interés pasó de la neurología a la etología (me gusta pensar que pasé de ser un reduccionista a ser un holista) gracias al estudio de la conducta de perros, pájaros, roedores y focas. Sin embargo, mi interés siguió centrado en los primates, por lo que acepté una ayuda postdoctoral para trabajar en el Oregon Primate Center en 1969, centro donde me he dedicado al estudio de los macacos japoneses desde 1970".

SHERWOOD B. IDSO ("Tempestades de polvo") es físico investigador del Departamento de Agricultura. Obtuvo el grado de B.Sc. en Física en la Universidad de Minnessota en 1964, donde continuó hasta obtener el Doctorado en 1967; se especializó en física de suelos y en meteorología e ingeniería mecánica. Esa amplia base ha capacitado a Idso para publicar más de cien trabajos científicos sobre temas que van desde micrometeorología básica hasta la influencia y (ausencia de influencia) del hombre sobre el clima global. "Me he especializado en tempestades de polvo y trombas. El lugar donde vivo es ideal en cuanto a observaciones de primera mano.

MANUEL BALLESTER ("La química percloroorgánica") dirige el Instituto de Química orgánica de Barcelona, del Consejo Superior de Investigaciones Científicas. En esa misma ciudad, en donde nació en 1919, cursó la carrera de Químicas, obteniendo el grado de doctor por la Universidad de Madrid en 1948. Fue becario de investigación química en la Universidad de Harvard (1949-1951) y profesor visitante en la Ohio State University and Aerospace Research Laboratories, Wright-Patterson AFB, Ohio, a cuyos laboratorios volvería años más tarde como director de los programas de investigación. Miembro de numerosas asociaciones de químicos (The Association of Harvard Chemists, The American Chemical Society y Real Sociedad Española de Física y Química, entre otras), ha visitado como profesor invitado varias universidades norteamericanas y europeas. En este trabajo ha contado con el patrocinio económico de los Aerospace Research Laboratories y del III Plan de Desarrollo Económico y Social español y con la colaboración de los profesores J. Castañer y J. Riera.

La erradicación de la viruela

Una campaña de 10 años de duración, dirigida por la Organización Mundial de la Salud, está alcanzando ya su objetivo. El mundo quizás ha presenciado el último caso de esta terrible enfermedad, la más devastadora de la historia

Donald A. Henderson

En dos apartadas regiones de Etiopía, en el desierto y las montañas respectivamente, se está librando la postrera batalla contra la viruela, la enfermedad epidémica más devastadora y temida a lo largo de la historia de la humanidad. En efecto, a finales de agosto, sólo cinco pueblos registraron casos en el período que cubría las ocho semanas últimas. Y lo que es más importante, el último caso conocido se declaró el 9 de agosto. Al ser el hombre la única especie en la que puede medrar el virus de la viruela, la enfermedad podría quedar eliminada para siempre en cuanto se produjera la curación de la última persona infectada. Actualmente, un ejército de más de 1000 técnicos y auxiliares de sanidad etíopes, al que se han incorporado 10 epidemiólogos de la Organización Mundial de la Salud, está recorriendo palmo a palmo todo el país para asegurarse de que no existe ningún caso más. Si descubrieran otro, el paciente sería sometido a un aislamiento bajo vigilancia durante las 24 horas del día, vacunándose a todo aquel que hubiera estado en contacto con el infectado. Se intentaría con todos los medios seguir la pista de la infección hasta llegar al brote de la enfermedad, conocido y eliminado anteriormente. A partir del último caso, y por espacio de dos años, la búsqueda de nuevos brotes continuará. Si no se encuentra ninguno más, y si la comisión internacional de la OMS comprueba que la búsqueda ha sido exhaustiva, se declarará a la viruela erradicada de la faz de la tierra. Y se habrá cumplido el primer objeto de esta índole en la historia de la medicina.

La Asamblea Mundial de la Salud (organismo de control de la OMS) votó la campaña mundial, de 10 años de duración, en pro de la interrupción de la



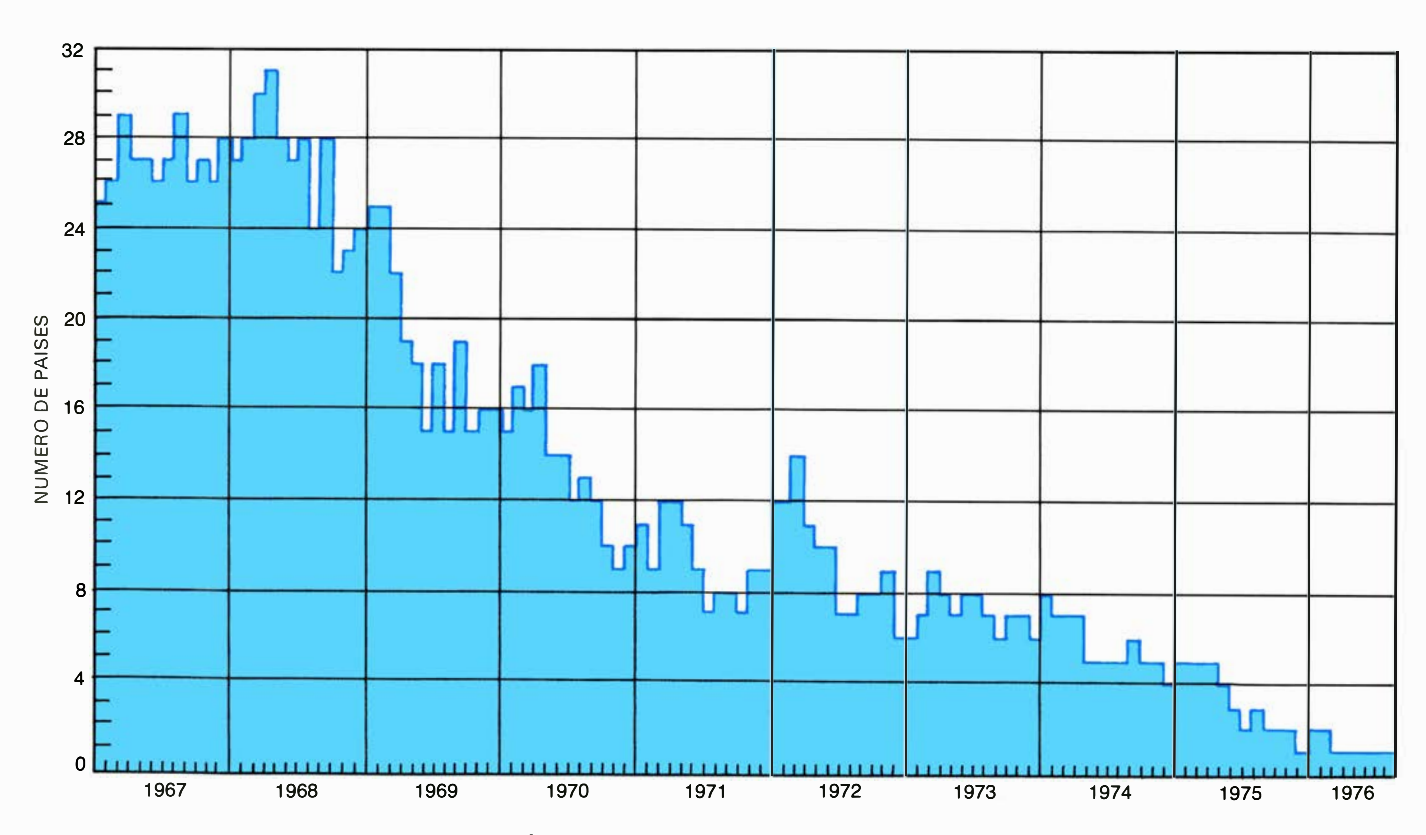
LA BUSQUEDA DE CASOS DE VIRUELA se lleva adelante en las remotas montañas de Simyen, en la provincia etíope de Begemdir. Se forman equipos de dos hombres, y se les asigna un área que deben recorrer a pie o en mulo con objeto de descubrir posibles casos nuevos y vacunar a todas las personas que pudieran haberse contagiado. El personal supervisor (en este caso un epidemiólogo de la Organización Mundial de la Salud y un colega etíope) se mantiene en estrecho contacto con los equipos desplazándose a cualquier lugar mediante un helicóptero.

transmisión de la viruela; tal programa, puesto en marcha en 1967, tenía que terminar, prefijadamente, en 1976. En el momento de iniciarse la campaña, la viruela se consideraba endémica –una enfermedad indígena, presente desde siempre– en más de 30 países; regularmente llegaban noticias de casos "importados" en unos doce países más, cada año. El programa fue avanzando de una forma irreversible a través de enérgicas campañas que fueron eliminando la viruela, sucesivamente, del Africa occidental y central, Brasil, Indonesia, Africa del Sur, Pakistán, la India y Bangladesh. En Etiopía, último de los países infectados, la campaña encontró mayores complicaciones por tratarse de una de las regiones más abruptas e inaccesibles de Africa; se calcula que más de la mitad de sus 28 millones de habitantes residen a más de un día de camino a pie de cualquier carretera. Las luchas entre fuerzas del gobierno y grupos disidentes constituyeron un grave problema en repetidas ocasiones. Dos sanitarios etíopes murieron alcanzados por las balas, y los equipos de búsqueda de casos infecciosos tuvieron que retirarse en una ocasión de algunos distritos durante semanas. A partir de 1971, cuando empezó la campaña en Etiopía, la extensión de las áreas infectadas y el número de casos quedaron rápidamente reducidos, y hoy el éxito final parece encontrarse próximo.

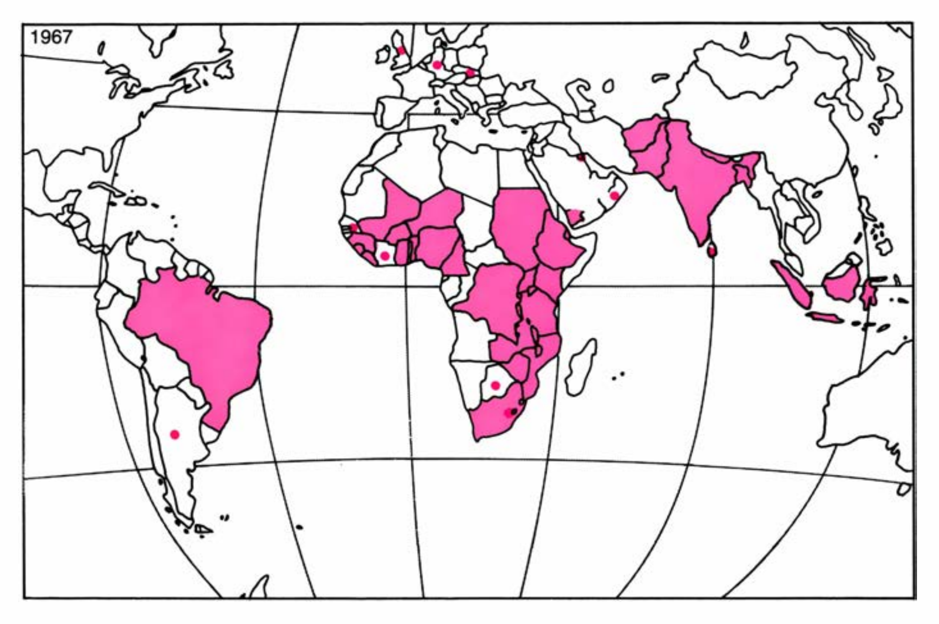
La viruela está provocada por un virus que se transmite por contagio de un individuo a otro a través de diminutas gotitas manantes por la boca o por la nariz. (Alguna vez, las partículas víricas que se encontraban en el vestido o en la cama han infectado a personas que no tuvieron contacto directo con el paciente pero esta infección indirecta fue siempre muy poco frecuente.) A los 10 o 12 días de inhalar el virus, la persona infectada siente malestar general, con fiebre elevada y sensaciones dolorosas que recuerdan las de una fuerte gripe. Dos o cuatro días más tarde, aparece una erupción en la cara, que al siguiente o dos días después se va extendiendo por todo el cuerpo. Suele manifestar una distribución "centrífuga": es más densa en la cara, brazos y piernas y menos densa en el tronco. Las pústulas que se originan, pequeñas, rojas, parecidas a granos, aumentan rápidamente de tamaño convirtiéndose en vesículas, llenas primero de una serosidad límpida y después, hacia el quinto día de la erupción, purulentas. En los casos graves, las pústulas pueden disponerse tan densamente, en especial sobre la cara y los párpados, que no quede piel normal; la cara aparece hinchada, y el paciente, ya gravemente enfermo, resulta irreconocible. Hacia el décimo día empiezan a formarse costras, que se van desprendiendo a las tres semanas, dejando zonas despigmentadas que se transforman en cicatrices cóncavas, desfiguradoras. Algunos pacientes acaban ciegos. Entre las personas afectadas por la forma asiática, virulenta, del virus, la *Variola major*, mueren de un 20 a un 40 por ciento. En Etiopía, en donde prevalece la *V. minor*, menos virulenta, la tasa de mortandad es del 1 por ciento. Una vez contraída la viruela, no existe tratamiento eficaz contra la misma.

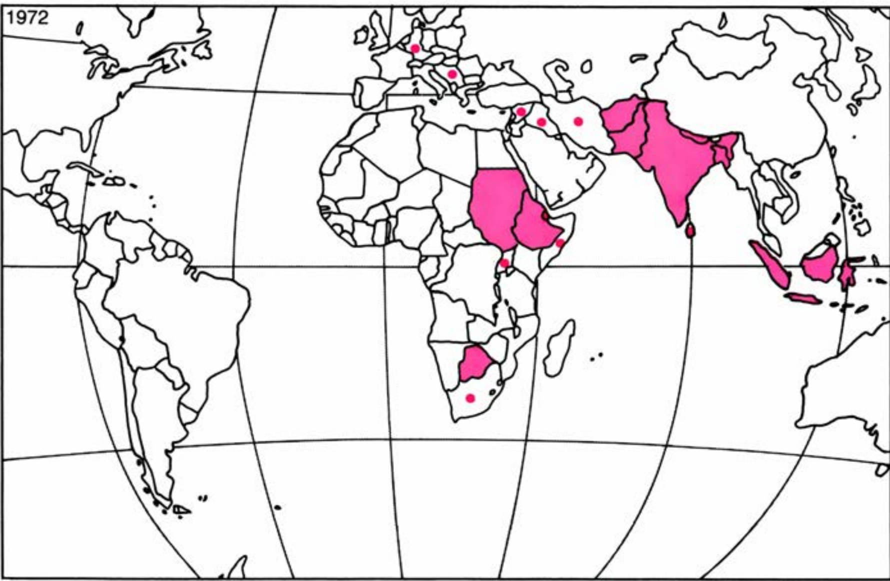
El origen de la viruela es anterior a la historia escrita. La cabeza momificada del faraón egipcio Ramsés V, que murió hacia el año 1160 a. C., a causa de una infección aguda, presenta lesiones que parecen corresponder a las producidas por la viruela. Los textos chinos y sánscritos indican que la viruela estaba también presente, en una época coetánea a aquél por lo menos, en China y en la India. En cambio, no existe mención alguna de una enfermedad que se parezca a la viruela ni en el Viejo Testamento ni en el Nuevo, ni en la literatura griega o latina, y cabe esperar que una enfermedad tan grave hubiera sido descrita, casi con certeza, de haberse manifestado comúnmente.

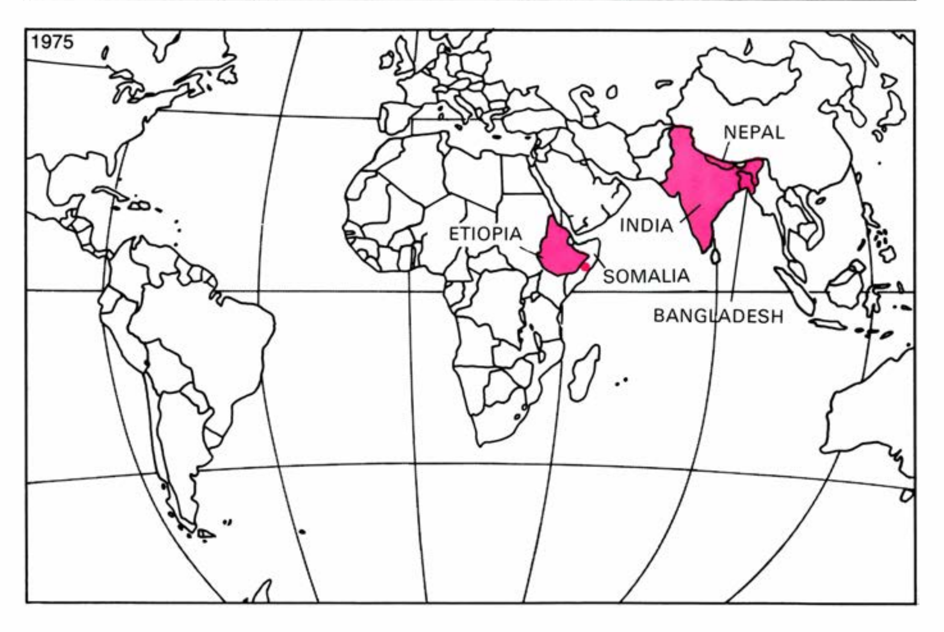
Una explicación plausible de ese fe-



LA VARIACION EN LOS ULTIMOS DIEZ AÑOS de la incidencia de la viruela puede seguirse en este histograma: en él figura el número de países con uno o más casos por mes. Se engloban los casos declarados endémicos y los "importados" en países donde no lo era.







EL AREA AFECTADA POR LA VIRUELA se ha ido reduciendo, como se desprende de los mapas correspondientes a 1967 *(arriba),* **1972** *(centro)* **y 1975** *(abajo).* **Los países en los que la enfermedad se consideraba endémica aparecen íntegramente coloreados; aquellos otros países que registraban casos importados están señalados mediante un punto o topo de color.**

nómeno se encuentra en las mismas características epidemiológicas de la viruela, las cuales justifican que sea posible eliminarla. Puesto que no existe ningún animal superior ni insecto conocido en los que el virus pueda refugiarse, para que la infección pueda persistir en una población es necesario que una persona atacada por el virus lo transmita a otro individuo sensible con quien haya establecido contacto; este contagio permitirá a su vez, transmitir la enfermedad a otra persona en una cadena ininterrumpida. La víctima de la viruela sólo puede transmitir la enfermedad desde el momento en que aparece la erupción hasta que las costras se desprenden, lo que corresponde a un período de unas cuatro semanas. En adelante, se vuelve completamente inmune a una nueva infección. En poblaciones aisladas y en núcleos de población dispersos se llega a un punto en el que quedan tan pocas personas susceptibles a la infección que se rompe la cadena de transmisión. En regiones de este tipo, la viruela desaparece, y no vuelve a aparecer salvo en el caso de que sea introducida de nuevo.

Parece razonable suponer que en la Antigüedad sólo las regiones de la India y de China, más densamente pobladas, eran capaces de mantener una transmisión continua de la viruela. Si ésta se introdujo de forma fortuita en Asia occidental, Europa y Africa, menos pobladas, pudo persistir durante cierto tiempo en una región como el delta del Nilo, pero, al final, la cadena de transmisón acabaría por romperse. La lentitud de los desplazamientos antes de la era cristiana y la escasa frecuencia de los largos viajes podrían haber servido para que la viruela quedara confinada, casi por completo, en Asia en ese período. Sin embargo, a principios de nuestra era en Asia occidental y a partir del siglo VI en Europa y Africa, se van haciendo más numerosos los relatos históricos que describen una enfermedad con todos los visos de ser la viruela. En 1520, los Conquistadores llevaron la enfermedad a América.

Al ir incrementándose la densidad de las poblaciones fue posible la existencia de suficientes individuos aptos para constituir los eslabones de la cadena de transmisión. Tan generalizada fue la difusión de la enfermedad que, al igual que el sarampión en nuestro tiempo, prácticamente todos la contraían. Los archivos de las parroquias inglesas de principios del siglo XVIII ponen de manifiesto que casi un 20 por ciento de las víctimas de la viruela morían, pero que la tasa de mortandad subió con ocasión de múltiples epidemias.

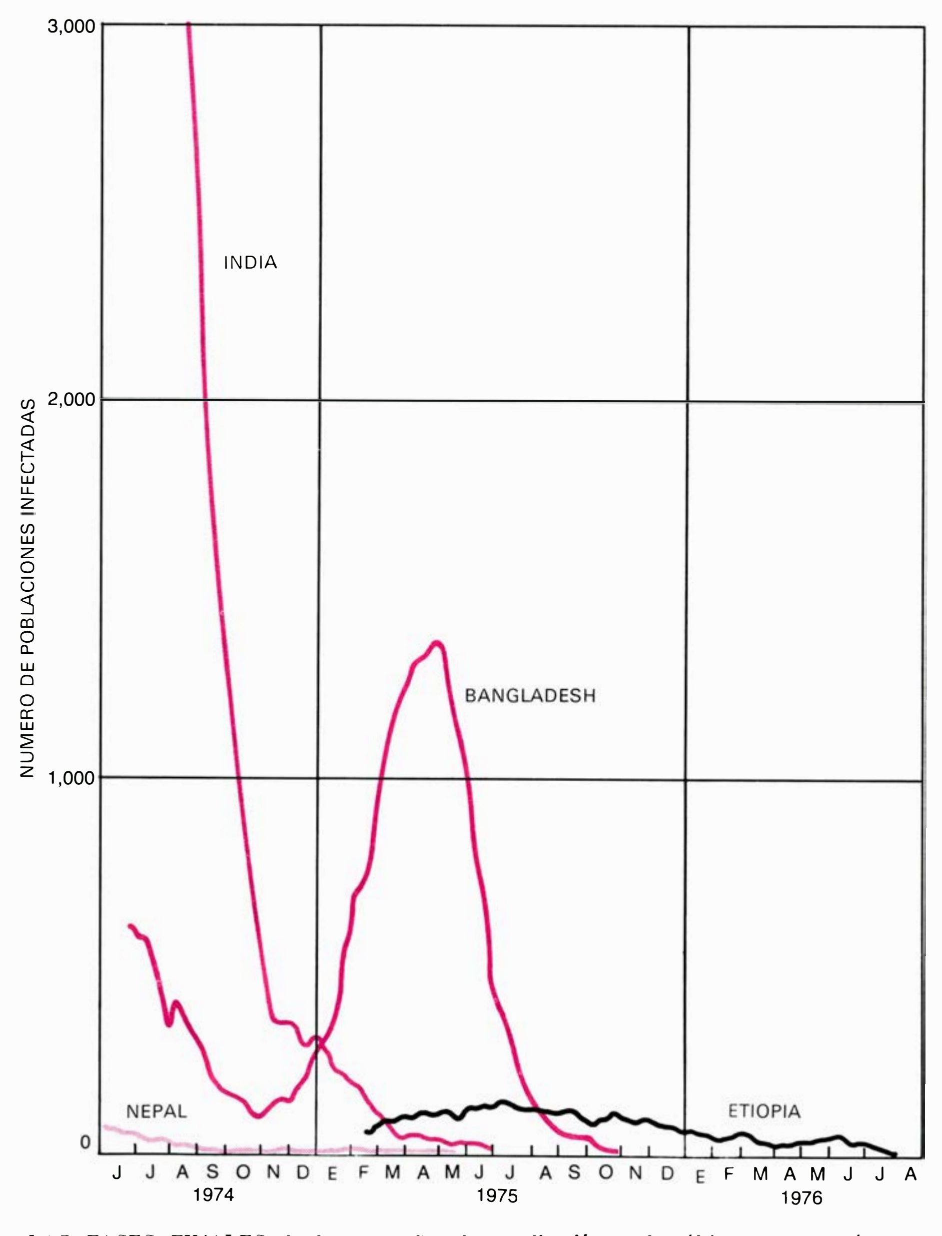
En un ejemplo bien documentado correspondiente a Islandia, tomado de 1707, 18.000 personas, es decir, un 31 por ciento de una población de 57.000 murieron a causa de la viruela. Si tenemos en cuenta que parte de la población era inmune por haber pasado ya tal enfermedad, la tasa de mortandad atribuible a esa epidemia pudo acercarse al 50 por ciento. Se cree que, en México, murieron por la viruela 3,5 millones de indígenas, poco después de haber sido introducida en el siglo XVI. Hasta el advenimiento de la vacunación, la viruela desempeñó un papel de primera importancia como freno del crecimiento demográfico.

Antes incluso de Edward Jenner, algunas personas habían sido protegidas contra la viruela mediante la variolización: materiales extraídos de la pústula de un paciente se incorporaban, por escarificación, en la piel de una persona sana, la cual, si todo iba bien, presentaba la forma benigna de la enfermedad y quedaba así inmunizada. La variolización se extendió desde Asia hasta Europa, pasando por el Próximo Oriente, como práctica popular; a principios del siglo XVIII, se adoptó y popularizó entre la clase médica inglesa. La tasa de mortandad a consecuencia de la variolización era sólo de un décimo de la que se alcanzaba a causa de la enfermedad adquirida naturalmente, pero una persona variolizada podía transmitir la viruela virulenta a otras. (Buscando los antecedentes de los brotes recientes registrados en Etiopía, se han detectado casos de esta práctica.) Cuando la variolización se propagó por Inglaterra, era ya algo conocido de todos que los ordeñadores que habían contraído una enfermedad llamada cow-pox no les "atacaba la viruela" al ser variolizados, y por tanto también resultaban inmunes a la infección de viruela natural. En 1796, Jenner inoculó material extraído de lesiones de cow-pox de una lechera en el brazo de un niño de 8 años de edad, a consecuencia de lo cual quedó inmune a los efectos de la variolización antivariólica. Jenner denominó al nuevo procedimiento vacunación, y predijo que "la aniquilación de la viruela será el resultado final de esta práctica". La nueva técnica se extendió rápidamente, pero sólo ahora, como resultado de un esfuerzo internacional concertado en el campo de la salud pública, la predicción de Jenner está recibiendo su cumplimiento.

Hasta finales del siglo XIX, el virus de la vacuna se obtenía a partir de las lesiones pustulares de una persona vacunada y se aplicaba al brazo de los demás individuos que se vacunaban por escarificación. Se trataba de un sistema muy poco eficaz, y no era infrecuente que la espiroqueta de la sífilis y el virus de la hepatitis fueran transmitidos al mismo tiempo que el virus de la vacuna. El descubrimiento de que se podía obtener gran cantidad del virus de la vacuna mediante escarificación e inoculación de la piel de un ternero, previamente afeitado, constituyó un importante paso adelante. (Este procedimiento continúa empleándose para obtener cantidades suficientes de virus para la producción de vacuna.) Sin embargo, la conservación del virus resultó difícil, en especial en climas tropicales. Algunas veces, se arrastraba el ternero de puerta a puerta, y se le extraía por escarificación una pequeña cantidad del pus del animal en cada vacunación.

Por regla general, el material de las pústulas raspado de la piel del ternero se dispersaba en una solución al 50 por ciento de glicerina, a fin de reducir la contaminación bacteriana, y se distribuía en pequeños capilares de vidrio. De esta forma, y sin refrigeración, la vacuna mantenía su vigor durante unos pocos días.

El propio Jenner había observado que la vacuna desecada duraba más que la vacuna líquida, pero por mucho tiempo no se encontró la manera de producir vacuna deshidratada en cantidades apreciables. A punto de estallar la Segunda



LAS FASES FINALES de las campañas de erradicación en los últimos cuatro países en donde la viruela era endémica se fueron controlando mediante contaje semanal de los "pueblos infectados", entendiendo por tales los que habían registrado algún caso dentro de las últimas semanas. Los contajes empezaron a establecerse en Etiopía, a principios de 1975; seis meses después, en el verano de dicho año, se consideraba su presentación como un total nacional.



EN ETIOPIA, las 14 provincias estaban infectadas de viruela en 1971, año en que se inició la campaña de erradicación bajo la dirección de la OMS. A finales de 1974, sólo quedaban siete provincias infectadas *(tonalidad media e intensa).* **A finales de agosto del citado año, únicamente había algunos pueblos infectados** *(discos blancos)* **en dos provincias** *(en color intenso)* **ubicadas en los desfiladeros del Nilo Azul y, la otra, en el desierto de Ogaden.**

Guerra Mundial se consiguieron desarrollar los métodos para la liofilización comercial de preparados biólogicos, y, a principios de la década de 1950, Leslie H. Collier, del Instituto Lister de Inglaterra, diseñó una técnica para producir una vacuna antivariólica liofilizada, notablemente termoestable. Collier produjo una suspensión de virus parcialmente purificada por centrifugación diferencial y liofilizó el material en ampollas, encerrado en las cuales podía distribuirse. Se reconstituía mediante una solución de glicerina, antes de la vacunación. La mayoría de los lotes de vacuna de Collier seguían siendo eficaces después de haber sido almacenados durante dos años a la temperatura corporal.

La vacuna liofilizada eficaz constituyó un arma esencial para hacer posible la erradicación de la viruela.

En 1926, cuando la sección de la Salud de la Sociedad de Naciones empezó la publicación de un boletín semanal sobre la importancia de las enfermedades en todo el mundo, se incluyó la viruela entre las enfermedades a tener en cuenta en la elaboración de estas estadísticas, no sin la oposición de algunos países. Cuando el delegado japonés de la Conferencia Internacional de Sanidad defendió la viabilidad de incluir también la estadística de la viruela, el delegado suizo arguyó que "en realidad, la viruela no tiene lugar en una convención internacional. No es una enfermedad epidémica en el sentido estricto de la palabra; en efecto, es una enfermedad que existe en cualquier parte. Probablemente no existe ni un solo país en el cual pueda decirse que no se registren casos de viruela".

Al fundarse en 1948 la OMS como organismo especializado de las Naciones Unidas, se ocupó en un principio del problema de la viruela de forma cautelosa; la comisión interina había sostenido dos años antes que "hasta entonces se hacía impracticable incluso uniformizar las vacunas antivariólicas". Sin embargo, a finales de la década de 1940, la idea de la erradicación de enfermedades en general, y de la viruela en particular, iba adquiriendo un número cada vez mayor de partidarios, entre cuyos convictos más destacados se hallaba Fred L. Soper, director de lo que era entonces la Oficina panamericana de Sanidad y es actualmente la Oficina regional de la OMS para ambos hemisferios americanos. Dirigieron la atención general sobre la demostración, verdaderamente práctica, de la eliminación de la viruela de Norteamérica y Europa en la década de 1940 y también –un logro muy notable– su erradicación de unos cuantos países con servicios sanitarios menos desarrollados, como Filipinas y algunas naciones de América Central. En 1950, con aportación de asistencia técnica a cargo de la Oficina panamericana de Sanidad, se emprendió una campaña en pro de la erradicación de la viruela del continente americano. Los resultados obtenidos fueron alentadores: en 1959 la viruela había sido totalmente eliminada de los países americanos, con excepción de Argentina, Brasil, Colombia, Ecuador y Bolivia.

El mismo año, a iniciativa de la URSS, la Asamblea Mundial de la Salud hizo un llamamiento a favor de la erradicación global y vacunación contra la viruela "en los focos en donde existe la enfermedad". La resolución señalaba que "los fondos destinados al control de la viruela y a la vacunación antivariólica en todo el mundo superan los necesarios para la erradicación de dicha enfermedad de los focos en donde es endémica".

La OMS y la Fundación de las Naciones Unidas para la Infancia (UNICEF) prestaron su ayuda a varios países para promover centros productores de vacunas. De distintos países se recabaron donaciones de vacunas, y en algunos se pusieron en marcha programas de vacunación en masa. Pero el progreso fue lento. El interés y las energías de algunos gobiernos y de los equipos sanitarios se habían centrado en un programa a escala general para extirpar la malaria; ante el temor de sus resultados inciertos, muchos dudaron también de la posibilidad de que cualquier programa de erradicación de una enfermedad dada llegara a buen término. Muchos Estados se encontraron con falta de personal especializado, vehículos y otras ayudas indispensables que les imposibilitaba llevar a cabo las campañas. Los que consiguieron llevar a término el proyecto se vieron a menudo acosados por casos de viruela importados desde los países vecinos. Las donaciones de vacuna resultaron sumamente deficientes para poder satisfacer las necesidades, y no existía además ningún mecanismo que

hiciera posible un control regular de la cantidad de vacuna producida en los centros de producción de implantación reciente. La calidad de la vacuna descendió, y los países que se vieron administrando una vacuna de escasa calidad perdieron toda ilusión.

En 1966, los delegados de la Asamblea Mundial de la Salud llegaron a la conclusión de que si no se destinaba dinero suficiente para un programa totalmente coordinado, había que abandonar el propósito de una erradicación general de la viruela. Después de un prolongado debate, y a pesar de muchos recelos, los delegados votaron un presupuesto especial de 2,5 millones de dólares para un programa intensivo que empezaría el 1 de enero de 1967. El objetivo era la extirpación de la viruela en todo el mundo a finales de 1976.

En 1967, el año en que empezó el programa intensivo de erradicación, se consideraba que la viruela era endémica en 33 países, mientras que once más daban cuenta de casos debidos a importaciones. En América, Brasil era el único país en donde la enfermedad era endémica, pero él sólo constituía la mitad de un continente. En el resto del mundo existían tres refugios principales de irradiación de la viruela. Uno de ellos, en Asia, se extendía desde lo que es ahora Bangladesh, a través de la India, Nepal y Pakistán, hasta Afganistán. Un segundo refugio comprendía prácticamente todo Africa al sur del Sahara. La tercera área de refugio correspondía al archipiélago de Indonesia.

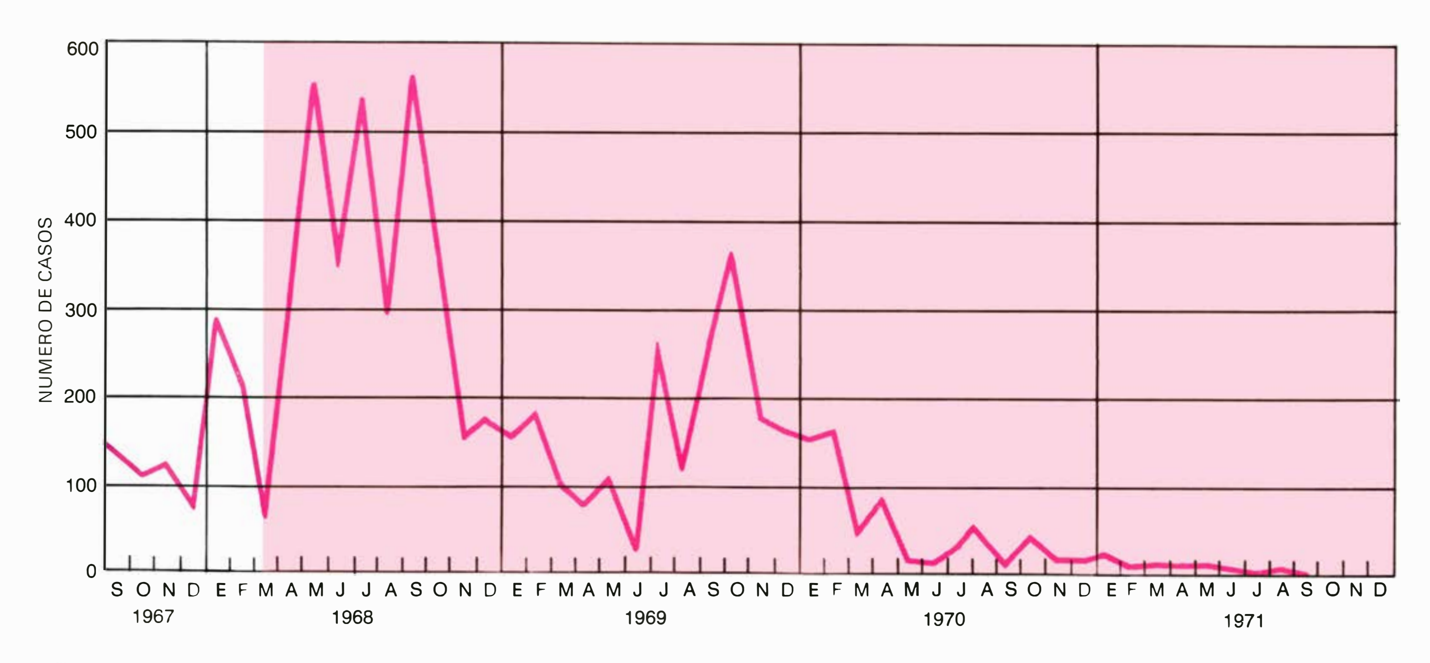
Al empezar el programa existía la urgente necesidad de suministros adecuados de vacuna eficaz y estable liofilizada. En el mundo se producía poca vacuna liofilizada de calidad satisfactoria. Algunas vacunas carecían de potencia cuando se reconstituían y otras perdían rápidamente su potencia en condiciones de campaña: algunas de las que estaban siendo administradas no contenían virus alguno detectable. Casi ninguna vacuna empleada en los países endémicos cumplía con las especificaciones básicas de la OMS y no existía ningún laboratorio centralizado que controlase la calidad de las vacunas.

A petición de la OMS, dos grandes laboratorios se incorporaron presto a la campaña actuando como centros de relación internacional para la vacuna: el Rijksinstituut voor de Volksgezondheid, en Holanda, y los Connaught Laboratories, Ltd., de Toronto. Se calculó que para las regiones endémicas se necesitarían del orden de 250 millones de dosis de vacuna al año. La compra de esta elevada cantidad importaba un coste superior al asignado a la totalidad del programa, por lo que se decidió no comprarla sino recabarla en forma de donaciones, y dotar de equipo y asistencia técnica a los laboratorios de las regiones endémicas para que, por lo menos los países endémicos más poblados, pudieran producir ellos mismos vacuna de calidad.

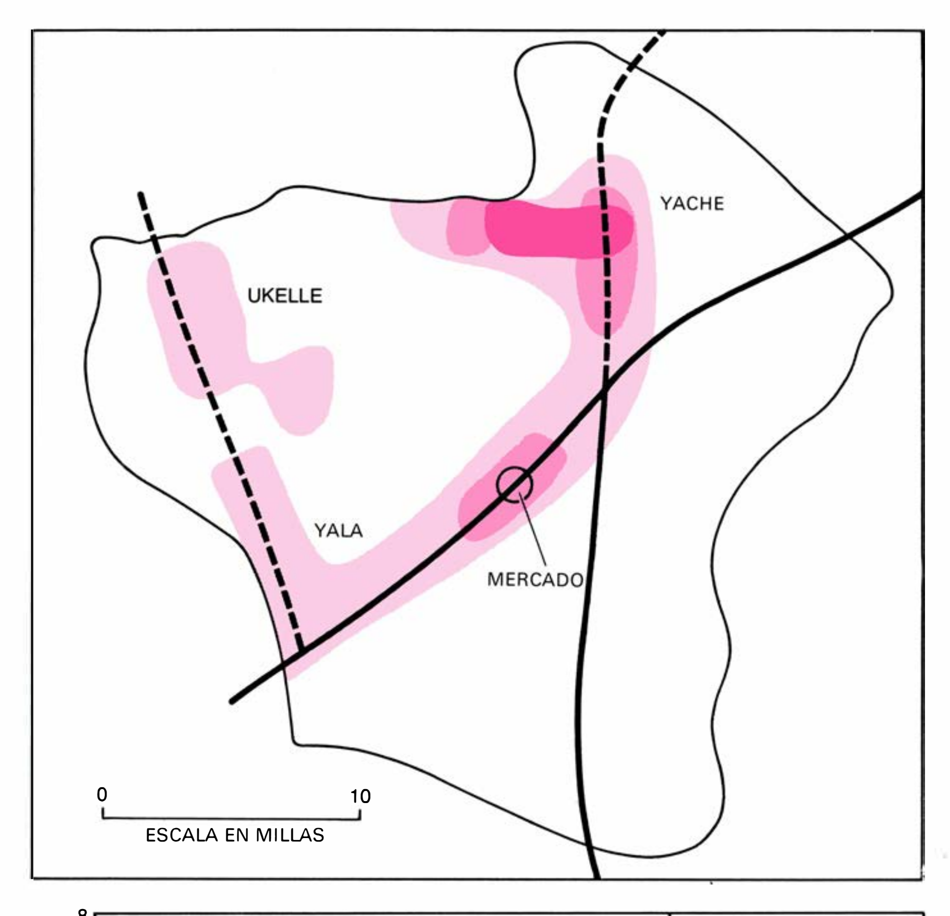
En los primeros años, el mayor volumen de vacuna –más de 140 millones de dosis al año– lo entregó la URSS; Estados Unidos aportó 40 millones de dosis. Las donaciones de vacuna se recibieron también de unos 20 países más. La producción de vacuna en las naciones en vías de desarrollo fue incrementando gradualmente, hasta conseguirse que en 1970 la totalidad de la vacuna empleada en el programa cumpliera con los niveles de eficacia y estabilidad aceptados internacionalmente.

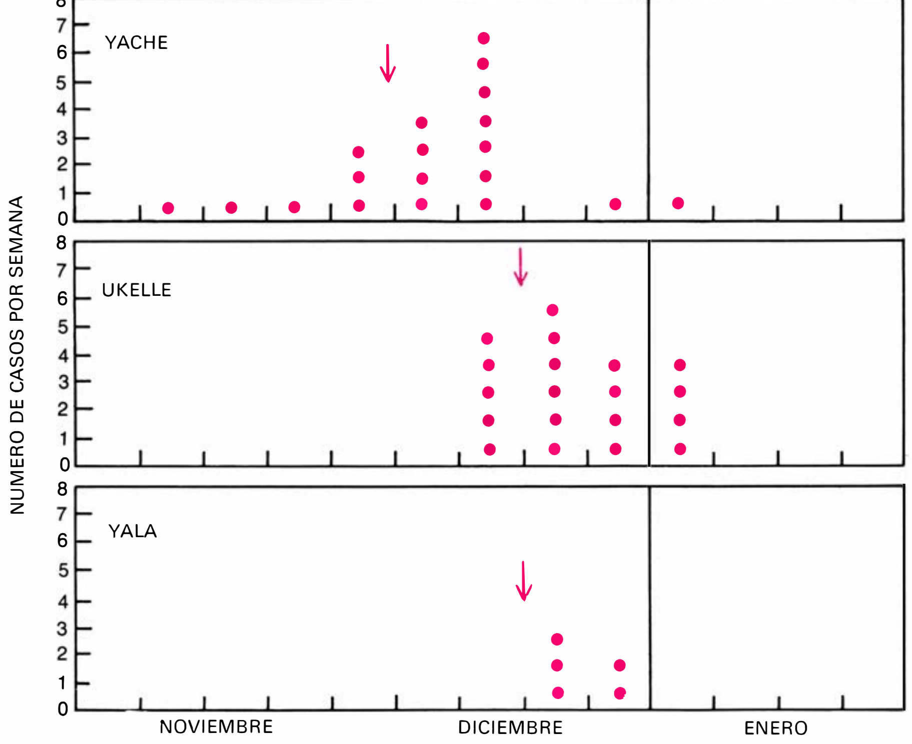
Otro objetivo inicial era la simplificación y la mejora de las técnicas de vacunación. Al principio del programa, la técnica más empleada en el mundo derivaba del método de la escarificación: se aplicaba sobre la piel una gota de vacuna y se escarificaban las capas superficiales, a veces con estiletes que producían heridas de consideración. Esta técnica desperdiciaba mucha vacuna y conseguía una proporción de éxitos más baja que el siguiente método: se mantenía un alfiler paralelo a la piel y se presionaba repetidamente la punta para que penetrara en la piel. Sin embargo, este procedimiento de presiones múltiples exigía mucha pericia y costaba aprenderlo. Había que encontrar un método más eficaz y que consumiera menos vacuna.

Un inyector a chorro, diseñado inicialmente para aplicar inyecciones subcutáneas más profundas, se modificó a fin de hacerlo práctico en vacunación antivariólica, más superficial. Se desa-



EN ZAIRE (denominado a la sazón República Democrática del Congo) se puso en marcha, en marzo de 1968, un programa intensivo de erradicación de la viruela. La curva de casos conocidos es típica: a medida que el programa de erradicación iba cumpliéndose, mejoraba la recogida de datos dimanantes de los casos de enfermedad, y esta fue la causa de lo que parecía un aumento en la incidencia de la viruela. Seis meses después empezó a producirse un descenso bastante constante, aunque irregular, y se tomó nota del último caso en agosto de 1971.





UN BROTE PRODUCIDO EN NIGERIA al principio de la campaña se trató mediante la técnica de la contención, cuya elevada eficacia motivó que más tarde substituyera ampliamente la inmunización en masa en Asia. El brote se detectó el 4 de diciembre de 1966, en Yache, población del distrito de Ogoja *(véase el mapa de arriba).* Este mismo día se empezó a aplicar la vacunación de la zona *(color intenso)* que rodeaba los casos conocidos; a medida que se iba teniendo noticia de nuevos casos, se extendió a las zonas afectadas *(color medio y claro).* La tabla *(abajo)* da cuenta de los casos semanales aparecidos hasta la inhibición de la vacunación. En cada pueblo se logró que el brote se extinguiera entre las dos y las tres semanas después del comienzo de la vacunación *(trazos),* con la excepción de dos individuos de Yache, que no habían quedado bien inmunizados. El experimento puso de manifiesto que era posible extinguir un brote infeccioso con la exclusiva vacunación de las personas que vivían en una zona limitada alrededor de cada nuevo caso, aun cuando en el resto del área vivieran muchas personas que no hubieran sido vacunadas.

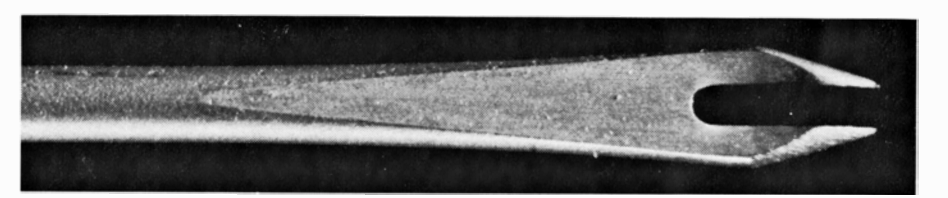
rrolló en el ejército norteamericano. Mediante fuerte presión, inyectaba una décima de mililitro de vacuna en las capas superficiales de la piel. Tal pistola inyectora requería una vacuna de mayor grado de pureza que los métodos de escarificación, pero podía contener menos de un tercio de la concentración en virus; podía vacunarse a más de 1000 personas (en el caso de que fuera posible reunirlas) por hora. Empleada, en 1967, en las campañas de Brasil y Zaire y en otros 20 países de Africa occidental y central, con la asistencia del Center for Disease Control, de Estados Unidos, las pistolas inyectoras demostraron su eficacia, pero eran caras y requerían considerables gastos de mantenimiento y reparación, por lo que no tenían sentido en los lugares carentes de expertos. Tampoco podían aplicarse en Asia, en donde la vacunación se hacía tradicionalmente de casa en casa.

La solución final, la aguja bifurcada, se desarrolló a lo largo de la década de los sesenta por los Laboratorios Wyeth. Se parece a una aguja de coser gruesa, de punta roma, con el ojo cortado y afilado por su parte apical con objeto de convertirlo en dos pequeños dientes. El extremo bifurcado de la aguja se sumerge en la vacuna; de esta forma, una gota de ésta, suficiente para una vacunación, queda retenida entre los dos dientes. La aguja se sostiene en posición perpendicular a la piel y con ella se practican 15 rápidas punciones. Esta técnica gasta sólo una cuarta parte de la vacuna que se requiere en la técnica de la escarificación. Wyeth renunció a los derechos de patente en los casos en que los fabricantes produjeran las agujas exclusivamente para el programa de la OMS. Al principio, esta aguja había sido diseñada para una vacunación, pero la OMS mandó fabricarlas con acero especial, de suerte que fuera posible hervirla o flamearla más de 200 veces sin que se alterara.

La técnica de vacunación ganó de nuevo en simplicidad cuando los estudios de campaña pusieron de manifiesto la inapreciable diferencia en la incidencia de infecciones bacterianas si la piel vacunada se había higienizado antes o no con acetona, alcohol o jabón. Ello permitió prescindir del algodón en rama y de las botellas de jabón y de alcohol. El vacunador necesita tan sólo un frasco de diluyente para la vacuna, un recipiente con agujas estériles y otro recipiente para las agujas usadas, todo lo cual cabe cómodamente en el bolsillo de la camisa.

Al principio del programa existían zonas donde la enfermedad era fuerte-

mente endémica, pero se desconocía en absoluto su relevancia y su extensión. Pronto se observó que la forma rutinaria de dar cuenta de los casos de viruela, en especial en las zonas en donde ésta era endémica, resultaba mucho peor de lo que se había supuesto; tan desastrosa mostró ser, que, hubo años en los que se registraron más muertes por viruela en grandes regiones que casos de tal enfermedad, de los que se hubiera tenido constancia. Un estudio realizado en el norte de Nigeria reveló que la relación entre casos habidos y casos de los que se hubiera dado cuenta era del 8,1 por ciento en zonas urbanas y del 1,3 por ciento en zonas rurales; en el oeste de Java sólo se había declarado un 6 por ciento de la totalidad de los casos. A la luz de la experiencia recogida en nueve años, hoy podemos sostener que, en realidad, al principio del programa sólo se tenía información del uno por ciento de todos los casos. Aunque en 1967 se dio cuenta a la OMS de 131.418 casos de viruela en todo el mundo, esta cifra debería corresponder únicamente a los casos habidos en el norte de Nigeria. Una estadística precisa para todo el mundo, para ese mismo año, podría haber arrojado perfectamente una cifra del orden de 10 a 15 millones de casos.



LA AGUJA BIFURCADA fue el principal instrumento empleado en el programa de vacunación. Este sencillo instrumento fue desarrollado por los Laboratorios Wieth. El extremo bidentado aparece aquí aumentado unas diez veces; su tamaño natural, es de unos 5 centimetros.

Había más. Constituyó un serio problema el que ni siquiera los casos detectados y denunciados por los equipos sanitarios se incluían necesariamente en las estadísticas nacionales. Cuando empezó la campaña en Indonesia, sólo la mitad aproximada de los casos denunciados a nivel provincial se recogía en las estadísticas a nivel nacional; en Nigeria, de 325 casos denunciados a los equipos del gobierno local, entre diciembre de 1966 y febrero de 1967, sólo 14 llegaron finalmente a conocimiento del Ministerio de Sanidad. Por ello, se dio un primer paso, en las campañas nacionales de erradicación de la viruela, al aumentar la calidad y la regularidad de la información sobre los casos aportados por cada equipo médico. Se intentó solucionar tal dificultad asignando un equipo de vigilancia, constituido por 2-4 personas, a cada división administrativa, que comprendía una población de dos a cinco millones. Estos equipos acudían a los centros médicos y hospitalarios estimulando la búsqueda y denuncia de casos, distribuyendo vacunas, aconsejando y animando al personal facultativo.

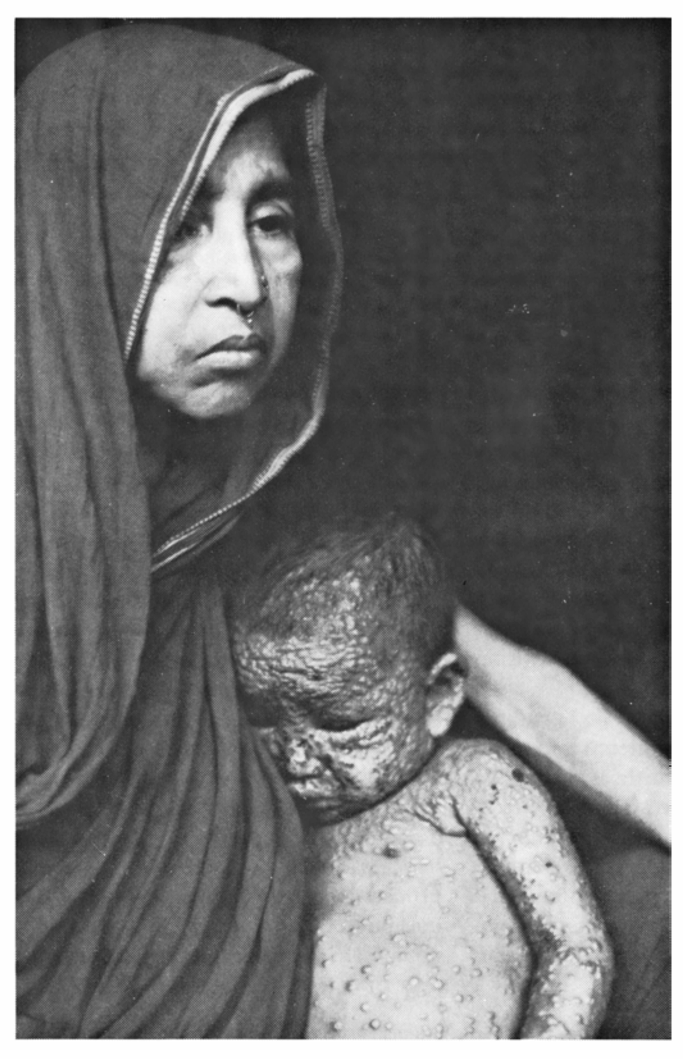
Cuando empezó el programa de erradicación intensiva, la táctica básica exigía una campaña de vacunación general, de dos a tres años de duración, que afectaba a todo el país, durante la cual se habría de ir desarrollando un sistema mejorado de localización de casos y de vigilancia. Se supuso que si el 80 por ciento de la población podía ser vacunado, la incidencia de la viruela se reduciría a menos de cinco casos por 100.000 habitantes. Luego, con una vigilancia más estricta, se irían identificando rápidamente los focos restantes, y así se podrían tomar las medidas oportunas hasta su total extirpación.

En los primeros tiempos de la campaña se hizo un descubrimiento importante que indujo a cambiar de estrategia. Al retrasarse el suministro de material necesario para la campaña de vacunación general en el este de Nigeria, un enérgico consejero estadounidense. William H. Foege, organizó un programa provisional: identificó casos de viruela y vacunó de forma exhaustiva a la población de un área limitada alrededor de cada caso *(véase la ilustración de la página opuesta).* El material para la campaña tardó contados meses en llegar, pero cuando lo hizo no quedaba viruela detectable en el este de Nigeria. Y se había vacunado casi la mitad de la población.

Este resultado y el de prácticas parecidas realizadas en otras regiones llevó a conceder la máxima importancia, desde las mismas primeras fases de la campaña, a lo que se denominó vigilancia-contención: búsqueda más eficaz y rápida detección, aislamiento de los pacientes y vacunación de cualquier persona de la que se sepa o se sospeche que ha entrado en contacto con ellos. Este procedimiento cortaba cualquier brote que pudiera producirse en el resto de la población. Se vio que tal sistema era



EL CARTEL, pegado a la pared, anuncia la recompensa para quienes declaren algún caso de viruela, en Bangladesh. El texto, en bengali, ofrece un "premio de 250 taka" (unos 17 dólares) a toda persona que denuncie un caso en la delegación de sanidad. Los empleados de sanidad se dedicaron a rastrear las informaciones, rumores incluidos, de posibles casos.





DOS PACIENTES DE VIRUELA, ambos de Bangladesh, aparecen con sus madres. En el caso grave, de la izquierda, las pústulas están dispuestas tan densamente sobre la cara que se juntan unas con otras; la erupción es menos densa sobre el tronco. A la derecha, esta niña de tres años, que se encuentra recuperándose de la viruela, corresponde a la última víctima conocida en Bangladesh, lo que la constituye también en el último caso de que se tenga conocimiento de la forma más virulenta de la enfermedad: la *Variola major,* de altísima mortandad.

parcialmente eficaz, en cierto modo por que en la mayoría de las regiones de viruela volvía a extenderse más lentamente de lo previsto. En general, una persona infectaba sólo a otras 2-5 más. En general, los casos aparecían agrupados en una parte de la ciudad o en pueblos adyacentes, y no se extendían amplia y uniformemente por el país. Los equipos de contención concentraron sus esfuerzos sobre áreas-problema y dispusieron así de más tiempo del calculado previamente, durante el cual podían cortar completamente un determinado brote. Cuando quedó patente que la viruela raramente aparecía entre los vacunados, incluso si lo estaban desde hacía muchos años, se concedió la máxima importancia a la primera vacunación (en Asia, el 80 por ciento total de los casos de viruela correspondió al 20 por ciento de las personas que no presentaban cicatriz de vacunación).

En los cuatro años subsiguientes a 1967 se realizaron importantes progresos. En 1971 sólo 16 estados declaraban casos de viruela. En los 20 países de Africa occidental y central que recibía ayuda del Center for Disease Control, de Estados Unidos, el último caso registrado (en Nigeria) data de junio de 1970, un año justo antes de lo esperado. En Africa, la viruela endémica quedó circunscrita a Etiopía, cuya campaña no empezó hasta 1971, a la parte meridional de Sudán, en donde la guerra civil había hecho imposible la aplicación del programa, y a Botswana, que volvía a reinfectarse por el tiempo en que se producían los últimos casos en la vecina Africa del Sur. El último caso de Brasil se declaró en abril de 1971, y el de Indonesia en enero de 1972. Quedaba una zona importante de refugio de la viruela. En Asia.

En 1972 podía dibujarse una amplia franja en donde la viruela seguía manteniéndose fuertemente endémica; partía de Bangladesh y se extendía por el norte de la India y de Nepal hasta Pakistán. En esta zona, la detección de los casos era defectuosa y los sistemas de declaración de los enfermos arcaicos; no se había comprendido aún la importancia de la vigilancia y la contención, y cuando se acometió la fase de contención se hizo en general de forma insuficiente. Tibio fue también el apoyo al programa dado por las autoridades sanitarias; tantos eran los esfuerzos para yugular la viruela que habían ido fallando una y otra vez a lo largo de los años, que ya había echado raíces la creencia fatalista de que la enfermedad era inevitable y su eliminación imposible.

Durante el bienio 1972-1973, un epidemiólogo de la OMS, que trabajaba en la India siguiendo el movimiento de la infección de un caso declarado en el estado meridional de Andhra Pradesh, recompuso los eslabones hasta llegar a un estado vecino, que se creía libre de viruela. En una zona del mismo descubrió una gran epidemia. La situación,

conocida por los técnicos sanitarios de la zona, se había mantenido encubierta. Se imponía una rápida contención, pero ello comportaba la inmediata detección de todos los casos en una comarca de dos millones de almas. Todos los equipos sanitarios disponibles fueron movilizados en una búsqueda casa por casa, que exigió dos semanas para su preparación y una para su ejecución. Sacó a luz numerosos casos antes desconocidos amén de los recientemente descubiertos. La vacunación de contención en torno a cada brote eliminó a las pocas semanas la viruela del distrito. Tal práctica permitió percatarse del peligro que entrañaba una información insuficiente de los casos de viruela, pero también puso de manifiesto el beneficio que podía sacarse de una búsqueda intensa así como la existencia de un nutrido cuerpo sanitario desplegado por toda la India que podía ser movilizado rápidamente. Se trazaron planes para repetir el mismo rastreo, exhaustivamente, en el subcontinente hindú.

En octubre de 1973, la campaña de erradicación de la viruela de la India inició la primera de una serie de prospecciones, de una semana de duración, que debían realizarse con periodicidad mensual, y en las que intervenían más de 100.000 sanitarios. Los resultados de las distintas prospecciones fueron siniestros. El estado de Uttar Pradesh, con una población de 90 millones de habitantes, había venido declarando en los últimos tiempos un número de casos que oscilaba entre 100 y 300 por semana. La primera prospección semanal en dicho estado registró más de 7000 casos. Esto no fue todo: cálculos realizados más tarde revelaron que sólo la mitad de los pueblos que constaban oficialmente como examinados lo habían sido en realidad.

En adelante, las técnicas de prospección de casos fueron mejorando progresivamente. Durante el tiempo que mediaba entre la última semana de prospección y la siguiente, los equipos especiales de vigilancia se dedicaban a preguntar por los mercados y escuelas a fin de descubrir rumores de casos. Se ofrecía un premio a toda persona que diera cuenta de un caso de viruela, así como al sanitario que registrara la declaración. Conforme la curva de la incidencia iba descendiendo, subía la cuantía del premio, acompañada siempre de una amplia publicidad. Se aplicaron con mayor rigor las medidas de contención. Se dispusieron libros especiales en donde figuraban todos los habitantes de los pueblos infectados y los que vivían en un radio de una milla. Guardias contratados en la misma localidad se mantenían apostados día y noche en la casa de los pacientes de viruela para impedir que salieran y comprobar que todos los visitantes estuvieran vacunados.

A medida que la detección y notificación de casos fue mejorando, el número de casos citados aumentó sensiblemente. En 1974 se registraron en la India más de 218.000 casos, el total más elevado desde 1958. Algunos periódicos lo conceptuaron como un desastre, y más de uno publicó una auténtica noticia necrológica del programa de erradicación global. Sin embargo, los que intervenían en el programa comprendieron que era precisamente la táctica de prospección la que constituía la clave de la erradicación de la viruela en el subcontinente.

En junio de 1974 se adoptó una nueva unidad de medida: el "pueblo infectado". Un pueblo (o un barrio en una ciudad) se incluía en la lista de infectados si se citaba en él un caso, por lo menos, de la enfermedad; permanecía en la lista durante los seis meses siguientes a la aparición del último caso y hasta que un epidemiólogo hubiera certificado (después de una búsqueda especial) que la transmisión se había interrumpido. La epidemia alcanzó su cota máxima en julio de 1974, mes en que se llegó a contabilizar más de 7000 pueblos infectados. Este número descendió irreversiblemente hasta que, en noviembre, se bajaba a la cifra de 350; a partir de entonces, sin embargo, el descenso se paró. Este fenómeno inquietó, toda vez que estaba empezando una nueva estación postmonzónica, con la consiguiente transmisión más rápida del virus. Un nuevo equipo de epidemiólogos se unió a la lucha, la contención se hizo más rígida y, en enero, la incidencia de la viruela volvía a decrecer. El 24 de mayo de 1975 se declaraba el último caso conocido de viruela en la India.

Afganistán había quedado libre en 1972 y la última noticia de su existencia en Pakistán viene fechada en octubre de 1974. Nepal, amenazado siempre por la plaga de casos importados de la India, registró su último enfermo en abril de 1975. En Bangladesh, la erradicación costó mucho más. Se echaron las campanas al vuelo cuando, en 1974, el número de pueblos infectados en el país descendió a 91. Lamentablemente, casi todas las zonas infectadas se encontraban en una región afectada por la hambruna que se declaró tras las inundaciones, las más devastadoras que se habían producido en las últimas décadas. Miles de refugiados infectados se dispersaron por todo Bangladesh, y a pesar de esfuerzos heroicos resultó imposible frenar la difusión de la enfermedad. (En dos días, en la plaza del mercado, un mendigo infectó a 52 personas de 18 pueblos distintos.) En enero de 1975 ocurrió otra catástrofe: los suburbios más humildes de Dacca, la capital, cayeron arrasados por las bombas, y unas 500.000 personas tuvieron que abandonar su lugar de residencia. Al extenderse por el país, algunos infectados también por la viruela, provocaron docenas de brotes satélites. A finales de abril, el número de pueblos infectados alcanzó una cota máxima de 1908. A partir de entonces, bajo la dirección del propio presidente, se puso en marcha un programa de emergencia nacional; el número de pueblos infectados decreció brusca y continuamente. El 16 de octubre de 1975, se anotaba la aparición de lo que sería el último caso de Bangladesh. La paciente Rahina Banu, de tres años de edad, se convertiría en el último caso registrado en todo el mundo de la forma grave de la viruela, la *Variola major*. Después, 12.000 sanitarios, bajo la supervisión de unos 100 epidemiólogos rastrearon una y otra vez, casa por casa, todo el país. No encontraron ningún caso, ni resulta verosímil que se descubra ningún otro, a pesar de lo cual la vigilancia continuará aún por espacio de dos años.

Existió cierta preocupación porque la viruela, después de haber sido declarada totalmente eliminada, pueda emerger de nuevo procedente de algún animal desconocido que haya actuado como refugio del virus, de virus que hubiera conservado una actividad latente en antiguas cicatrices, o a partir de cualquier otro origen. La prueba más impresionante en contra de la verosimilitud de esta reaparición está en que, durante nueve años, no se ha descubierto ningún caso más en las extensas regiones que se han liberado de ella, excepción hecha en casos importados de zonas endémicas conocidas. Aunque en algunas regiones de Africa declaradas libres de viruela se han detectado 20 casos de una enfermedad muy parecida a ella, lo que se ha aislado o identificado mediante anticuerpos en el suero sanguíneo de la víctima es un virus causante de la enfermedad; se trata del virus de la "viruela de los simios". Se desconoce qué animal hospeda al virus de la viruela de los simios; pero habrá que buscarlo probablemente entre los roedores, y no entre los monos. La enfermedad, aunque relacio-

nada con la viruela, está provocada por un virus distinto, y su capacidad de transmisión de una persona a otra resulta ser prácticamente nula.

Importa qué pueda decirse acerca del coste de este programa de 10 años de duración. Se han gastado, aproximadamente, unos 83 millones de dólares en asistencia internacional para el programa de erradicación de la viruela, desde 1967. Los propios países en donde la enfermedad era endémica aportaron casi el doble de esta cantidad, pero son pocos los que gastaron mucho más de lo que ya estaban destinando al control de la viruela. La cantidad total de dinero internacional en asistencia internacional apenas supera la mitad de la que en 1968 se calculó que correspondía al gasto anual destinado al control de la viruela sólo en Estados Unidos; las expensas mundiales en vacunación antivariólica y medidas de cuarentena oscilan entre mil y dos mil millones de dólares al año. Una vez erradicada la en--fermedad, la vacunación antivariólica deja ya de ser necesaria, y, con ella, los certificados internacionales de vacunación antivariólica. Aparte la reducción del sufrimiento humano, el ahorro ha amortizado ya muchas veces la pequeña inversión realizada.

La erradicación de la viruela representará un importante hito en la historia de la medicina. Se habrá demostrado lo que puede lograrse cuando los gobiernos de todo el mundo se unen en una organización internacional para lograr un objetivo común. Sin embargo, visto en perspectiva, este programa representa sólo un tímido primer paso hacia la consecución de un nivel tolerable de salud pública en todo el mundo.

Otro paso que hay que dar lógicamente a continuación es el de la aplicación de lo aprendido en este programa en otros que se propongan la inmunización para control de la difteria, tos ferina, tétanos, sarampión, poliomielitis y tuberculosis. Existen vacunas eficaces contra todas estas enfermedades, pero, fuera de los países desarrollados, se ha logrado un bajo nivel de inmunización. Hoy, varios de los países en vías de desarrollo, con la confianza que da el éxito logrado y reconociendo el beneficio que puede obtenerse con muy poco esfuerzo, se han embarcado en nuevos programas de inmunización. Mediante campañas de este tipo se empieza a constatar un cambio perceptible, si bien lejos todavía de ser el justo, de una medicina terapéutica para ricos a una medicina preventiva para todos.

Generación fotovoltaica de electricidad

Una forma de aprovechar la energía solar es convertirla en electricidad con células solares del tipo de las usadas en las naves espaciales. Tales células son caras, pero se están ensayando nuevos tipos de fabricación más económicos

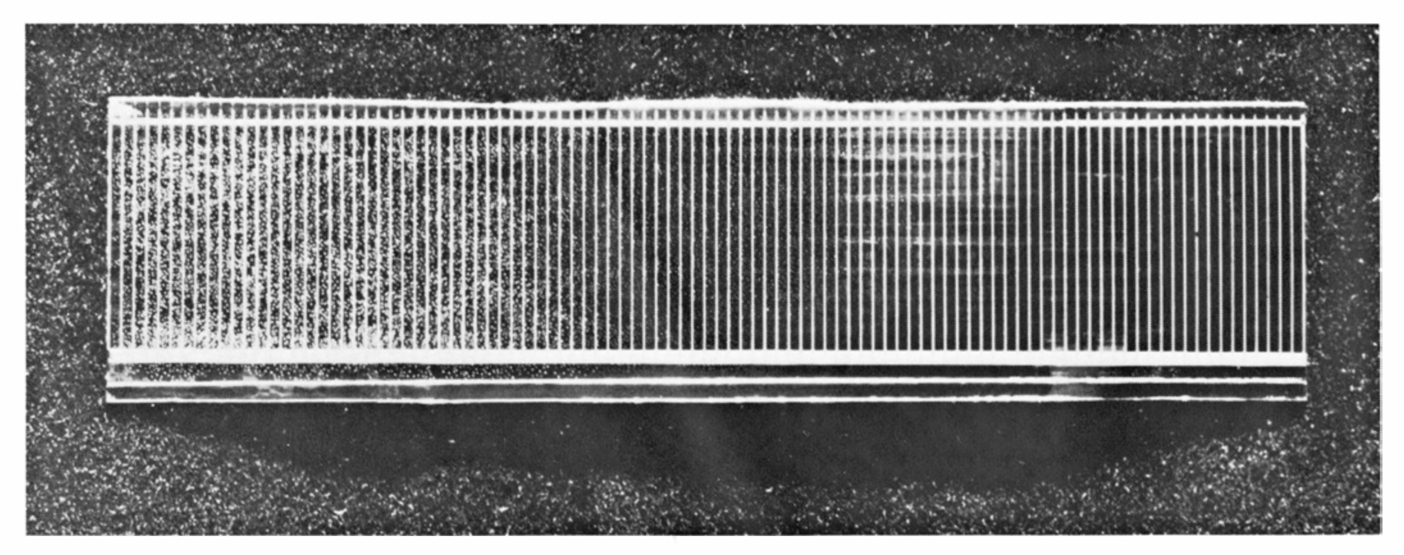
Bruce Chalmers

Resulta paradójico que a los americanos preocupe el suministro de energía cuando los Estados Unidos reciben energía del sol a un ritmo 500 veces superior al que la consumen; pero quizás la paradoja se pueda resolver. El hombre no ha hecho más que empezar a dirigir su ingenio científico y tecnológico hacia el objetivo de convertir su fuente de energía más abundante y segura, la luz solar, en electricidad, la forma de energía más conveniente y adaptable.

A pleno sol, en un cielo despejado, la energía irradiada alcanza la superficie de la tierra a una velocidad algo superior al kilovatio por metro cuadrado. Esta es una notable cantidad de energía; mientras el sol brilla, la energía equivalente a 4,5 litros de gasolina (36 kilovatios-hora) cae sobre un área del tamaño de una pista de tenis aproximadamente cada diez minutos; un área de sólo 80 metros cuadrados, en las partes menos soleadas de los 48 estados contiguos de la Unión, recibe a lo largo de un año la energía necesaria para satisfacer las necesidades de una familia media norteamericana. El sol es, también, una fuente de energía segura en el sentido de que en un lugar dado se recibirá aproximadamente la misma cantidad de energía solar cada año. Pero es una fuente insegura en el sentido de que la luz solar es intermitente: resulta inviable de noche o con tiempo desfavorable.

Antes de que la energía solar pueda proporcionar electricidad a gran escala hay que resolver dos problemas técnicos. El primero es el problema de convertir la energía de la radiación solar en electricidad. El segundo tiene que ver con el hecho de que la electricidad debe estar disponible cuando y donde se necesite, independientemente del tiempo, la estación del año y la hora del día. Hay que tener también en cuenta que en la realidad ambos objetivos se deben lograr a precios competitivos con los de otras fuentes de energía.

Desde hace más de cien años se sabe que la luz puede producir electricidad; una aplicación familiar de este conocimiento son los fotómetros empleados para medir la intensidad de la luz. Sin embargo, la eficacia de estos dispositivos es muy baja, y nunca podrán considerarse como medios prácticos de generar electricidad en cantidades útiles. El descubrimiento, en los años 50, de



CELULA SOLAR EXPERIMENTAL, hecha con silicio desarrollado por un nuevo proceso conocido como "crecimiento de láminas de bordes definidos" (edge-defined film-fed growth) (EFG) y fabricada por Mobil Tyco Solar Energy Corporation en Waltham, Mass. El proceso EFG es un método que promete disminuir los costes de construcción de las centrales fotovoltaicas a un nivel comparable al de la construcción de centrales térmicas. En particular, esta célula solar mide 10 cm de longitud y 2,5 de ancho. A pleno sol, puede generar entre un cuarto y un tercio de vatio de electricidad. Las líneas en blanco a intervalos regulares son los cables del colector de corriente que hay en la superficie frontal de la célula. Las motas finas que aparecen en la parte izquierda es polvo acumulado accidentalmente sobre la misma célula.

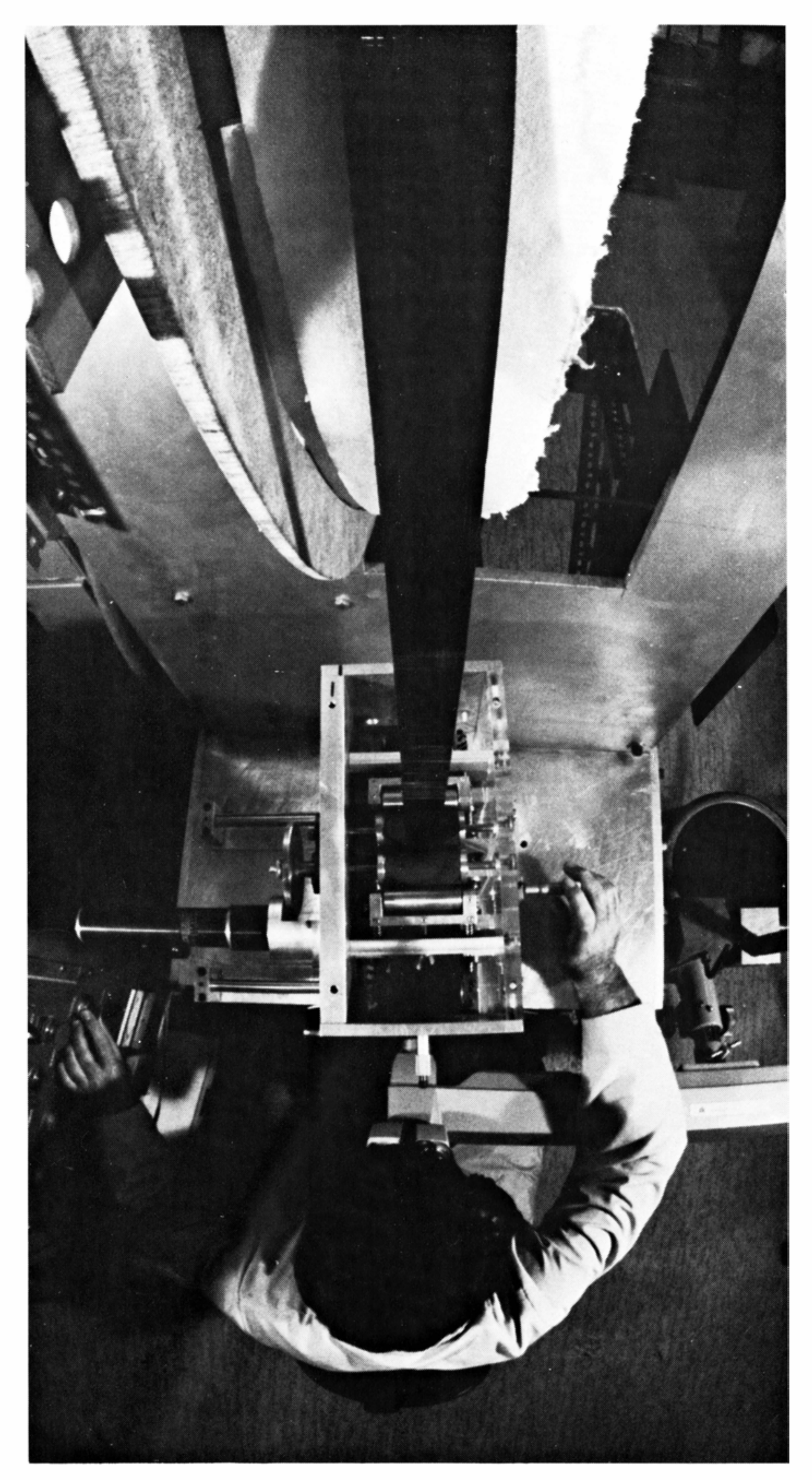
las extraordinarias propiedades eléctricas de un grupo de materiales que se dio en llamar semiconductores, llevaron al desarrollo de dispositivos en los que la eficacia de conversión de luz en electricidad es mucho mayor. Como ha ocurrido en otras muchas ramas de la tecnología, el programa espacial proporcionó el incentivo para el desarrollo y producción de los primeros sistemas fotovoltaicos prácticos, que han suministrado energía eléctrica en muchos vehículos espaciales, tripulados o no, con una fiabilidad muy alta y durante un tiempo aparentemente ilimitado.

Las células solares que se han empleado hasta ahora en el programa espacial han sido de silicio. Este material no es el único que se puede utilizar en dispositivos fotovoltaicos, y cabe que en el futuro se encuentre algún otro material superior al silicio. Sin embargo, ya que los principios en que se basan todos los dispositivos fotovoltaicos son los mismos, será conveniente describir su construcción y forma de operación en términos de la célula solar de silicio.

Una corriente eléctrica es sinónimo de un flujo de electrones en un circuito. Vamos a tratar, por tanto, del comportamiento de los electrones en un cristal de silicio. El silicio tiene valencia cuatro esto es, de los catorce electrones que tiene el átomo de silicio, cuatro son capaces de interacción con otros átomos. Pueden combinarse con átomos de otros elementos para formar compuestos químicos o pueden interaccionar con otros átomos de silicio para estabilizar la estructura del cristal.

Cuando los átomos de silicio pueden adoptar su configuración de más baja energía, como cuando el silicio fundido está cristalizando, cada átomo se rodea de cuatro vecinos equidistantes. Comparte dos electrones con cada vecino, uno suyo y otro del átomo vecino, y, por tanto, ocupa el centro de un tetraedro cuyos vértices están ocupados por otros átomos de silicio. El esquema anterior puede extenderse hasta el infinito, de forma que cada átomo, no solamente el central, está rodeado por otros cuatro átomos equidistantes. El resultado de esta disposición geométrica es la estructura cristalina cúbica característica del diamante.

Si la estructura del cristal de silicio fuese perfecta, cada electrón estaría fuertemente fijado en su posición por las fuerzas electrostáticas entre él y los dos átomos a los que mantiene ligados. Tal cristal sería un aislante perfecto ya que no habría electrones libres de moverse



CINTA DE CRISTAL DE SILICIO que crece en un baño de silicio fundido en el laboratorio de Mobil Tyco. La cinta tiene 2,5 cm de ancho y 0,15 mm de espesor, y emerge a través de la ranura de una cuña de grafito sobre la fundición a la velocidad de 2,5 cm por minuto. Para la fabricación de las células solares se corta en piezas de la misma longitud que la célula de la ilustración anterior. El cristal de silicio fabricado por el método EFG no es tan químicamente puro ni tan cristalográficamente perfecto como el crecido por el método Czochralski, usado para hacer cristales para dispositivos microelectrónicos. Resulta adecuado para fabricar células solares de gran rendimiento y más económicas.

si se aplicase un voltaje. En los metales, que son buenos conductores de la electricidad, los átomos también están unidos por electrones, suministrando cada átomo uno o más electrones. Sin embargo, en un metal, los electrones no están localizados de tal forma que enlacen pares específicos de átomos. Son libres de moverse dentro del cristal y dar lugar a una corriente eléctrica cuando se aplica un voltaje.

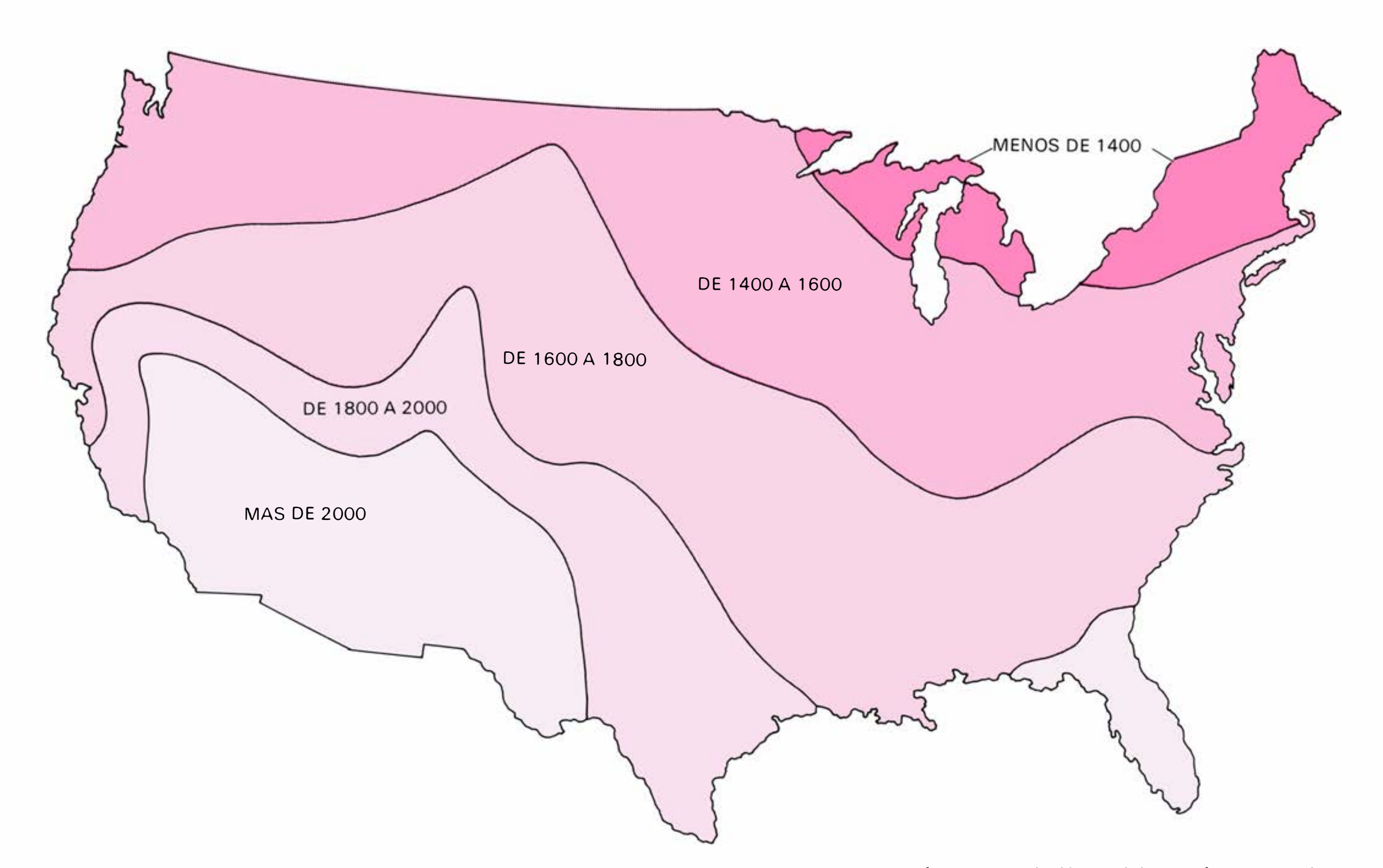
En un cristal de silicio se necesita una cantidad considerable de energía para romper el enlace entre un electrón y los dos átomos que éste une, para que el electrón pueda contribuir a la conducción eléctrica. Esta energía, que supone 1,1 electrón-voltios, puede suministrarse de varias formas. Se suministra espontáneamente siempre que un electrón reciba una cantidad suficientemente alta de energía térmica y, por tanto, a cualquier temperatura, excepto en el cero absoluto, siempre hay en el silicio algunos electrones disponibles para la conducción eléctrica, y el número de esos electrones aumenta con la temperatura. De esta forma, un cristal de silicio no es un aislante perfecto, sino que tiene una conductividad intrínseca. La conductividad es pequeña comparada con la de un metal pero es una propiedad importante del cristal.

Cuando un electrón alcanza el nivel de energía conocido como banda de conducción, donde puede participar en la conducción eléctrica, deja tras de sí un "hueco", un lugar donde falta un electrón. Un electrón ligado en las proximidades puede moverse hacia el hueco, ocupándolo, intercambiando sus posiciones. Por tanto, electrones y huecos pueden moverse dentro del cristal; y su movimiento conjunto constituye una corriente eléctrica. Cuando se aplica un voltaje, los electrones, que son la unidad básica de carga negativa, se mueven hacia el potencial positivo y, los huecos, hacia el negativo, como si estuviesen cargados positivamente.

Una segunda forma de proporcionar electrones o huecos para la conducción en un cristal es introducir en él átomos que suministren más o menos electrones que los cuatro que requiere la estructura ideal. Aquí, la energía que se necesita para crear huecos y electrones de conducción viene del desajuste eléctrico entre el átomo extraño y el cristal de silicio. Por ejemplo, el fósforo es un elemento cuya valencia es cinco, esto es, dispone de cinco electrones para interaccionar con otros átomos. Un cristal de silicio en el cual una pequeña proporción de las posiciones atómicas estén ocupadas por átomos de fósforo en lugar de átomos de silicio, seguirá teniendo la estructura básica del silicio. Sin embargo, habrá un número extra de electrones en la banda de conducción ya que no existen posiciones ligadas disponibles para ellos.

Un cristal de silicio dopado (impurificado) con fósforo es un semiconductor tipo-*n* porque tiene un exceso de electrones (exceso de carga negativa). Por el contrario, un cristal de silicio se puede dopar con boro, cuya valencia es tres; entonces el cristal es un semiconductor tipo-*p* porque tiene un exceso de huecos cargados positivamente.

Al dopar un cristal, no sólo se aumenta su conductividad sino que también se le convierte en un cristal preferentemente receptor de huecos o electrones. Un cristal tipo-*n*, que tiene ya más electrones de los que necesita para enlazar sus átomos (pero no más de los que necesita para su neutralidad eléctrica) puede inmediatamente absorber más electrones. Por el contrario, un cristal tipo-*p* tiene afinidad por los huecos.



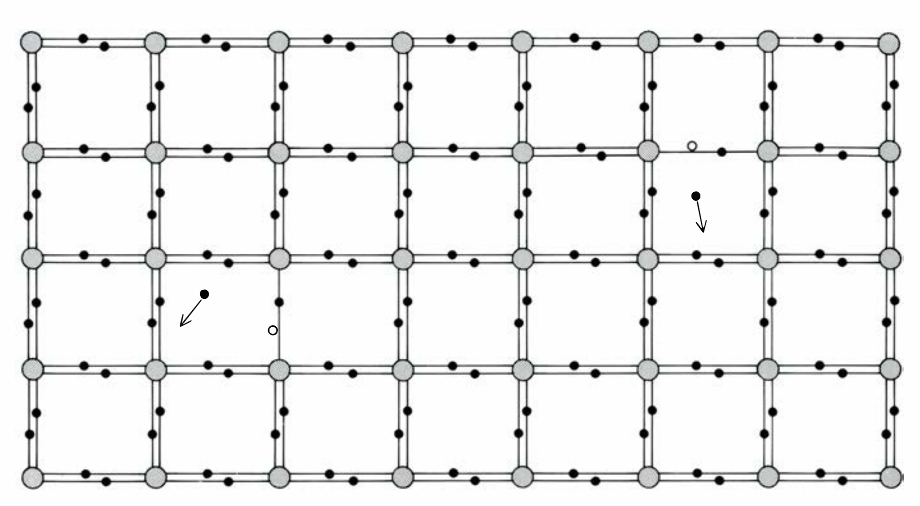
LA RADIACION SOLAR recibida cada año en diferentes zonas de los Estados Unidos viene dada en kilovatios-hora por metro cuadrado. La radiación recibida en los estados de Nueva Inglaterra es alrededor de dos tercios de la recibida en los estados del suroeste. Naturalmente, en cada área, la radiación recibida varía con el clima, la hora del día y la estación del año. Por esta razón, cualquier sistema de generación solar de electricidad a gran escala tendría que integrarse en sistemas convencionales de generación, o ser fácil de almacenar.

La luz incidente sobre el cristal es otra fuente de la energía que puede excitar electrones hasta el estado de conducción, creando electrones y huecos en igual número. Un fotón de luz que incide sobre un material es absorbido cuando su energía se transfiere a un electrón. Si el electrón absorbe la energía del fotón cerca de la superficie del metal, el electrón puede ser emitido desde la superficie; en esto consiste el conocido efecto fotoeléctrico. Si el fotón penetra en el semiconductor y su energía es igual o mayor que la energía que necesita para excitar el electrón hasta la banda de conducción, da lugar a un electrón de conducción y a un hueco.

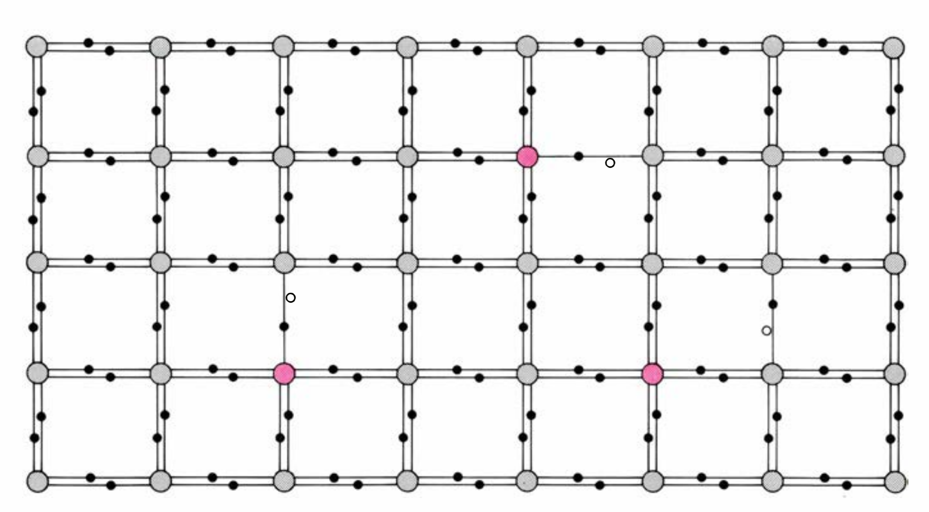
Por tanto, si luz de energía suficiente incide sobre un cristal perfecto de silicio se producen electrones y huecos capaces de moverse libremente dentro del cristal. Si al cristal no se le aplica potencial alguno, el electrón y el hueco se mueven al azar por el cristal. Cuando en sus viajes un electrón de conducción encuentra un hueco, "cae" en él. El electrón y el hueco se aniquilan mutuamente cuando el electrón cae de la banda de conducción y queda de nuevo ligado a la estructura del cristal. La energía que había sido absorbida para crear originalmente el par electrón-hueco se libera en forma de calor, aumentando ligeramente la temperatura del cristal.

Si se desea emplear electrones y huecos como fuente de electricidad, es necesario organizar las cosas de forma que un electrón no pueda recombinarse con un hueco hasta que haya viajado a través de un circuito externo, proporcionando un trabajo útil en su recorrido Esto se puede conseguir aprovechando las propiedades opuestas de los cristales tipo-*n* y tipo-*p*. Consideremos un cristal compuesto de una lámina de silicio tipo-*n* y una lámina de silicio tipo-*p*. Si se crean pares electrón-hueco cerca de la unión de los dos tipos de cristal, la afinidad del cristal tipo-*n* por los electrones y la del cristal tipo-*p* por los huecos reducirán lo ordenado de sus movimientos por el cristal, de forma que haya un flujo neto de electrones del cristal tipo-*p* al cristal tipo-*n* y un flujo neto de huecos en la dirección opuesta.

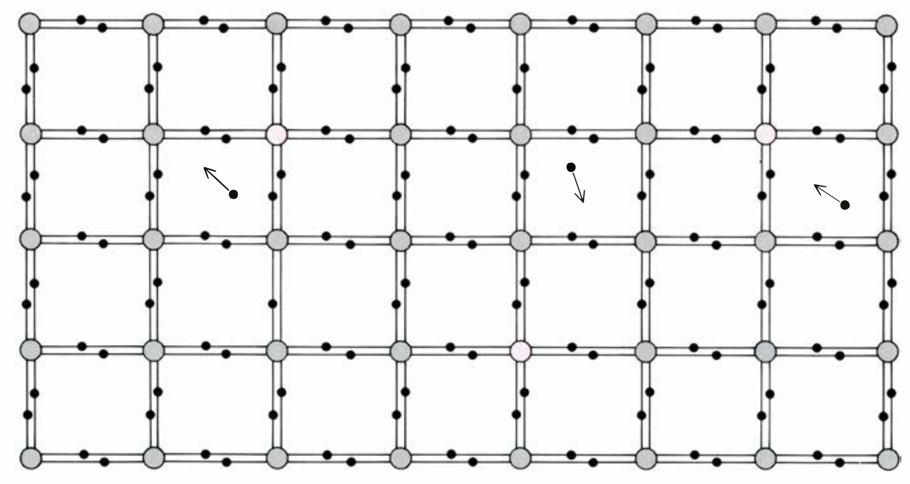
En una célula solar de silicio, la lámina de silicio tipo-*n* tiene un espesor tal que la luz que incide sobre su superficie penetra lo suficiente en el cristal como para crear pares electrón-hueco en las proximidades de la unión con el silicio tipo-*p*. Este espesor es del orden



UN CRISTAL DE SILICIO PURO tiene estructura cúbica; por mor de simplicidad, el de la figura está dibujado en dos dimensiones. El átomo de silicio *(gris)* tiene cuatro electrones de valencia. Cada átomo está fuertemente fijado en la red cristalina al compartir dos electrones *(negro)* con cada uno de sus cuatro vecinos más próximos. Posibles vibraciones térmicas o un fotón de luz proporcionarán suficiente energía como para hacer pasar un electrón al nivel de energía conocido como banda de conducción, donde el electrón se puede mover libremente por el cristal y es libre para conducir electricidad. Cuando el electrón abandone su anterior posición, dejará un hueco *(blanco)*, una región con carga neta local positiva.



UN CRISTAL DE SILICIO TIPO-*P* puede conseguirse dopando el sicilio con pequeñas cantidades de boro. Cada átomo de boro *(color)*, tiene sólo tres electrones de valencia, compartiendo dos electrones con tres de sus vecinos y uno con el cuarto. Este cristal tiene igual estructura que el silicio puro, pero más huecos que electrones de conducción.



UN CRISTAL DE SILICIO TIPO-*N* se obtiene dopando éste con pequeñas cantidades de fósforo. Cada átomo de fósforo *(color)* posee cinco electrones de valencia; así pues, no todos son parte de la red cristalina. De este modo, el tipo-*n* tiene exceso de electrones libres.

de media micra. Por consiguiente, cuando la luz incide en la célula, la lámina tipo-*n* recogerá electrones y la lámina tipo-*p* recogerá huecos hasta que se establezca un voltaje dentro del cristal suficiente para empujar cualquier electrón adicional hacia la lámina tipo-*p*. En la célula solar de silicio este voltaje es aproximadamente de 0,65 voltios.

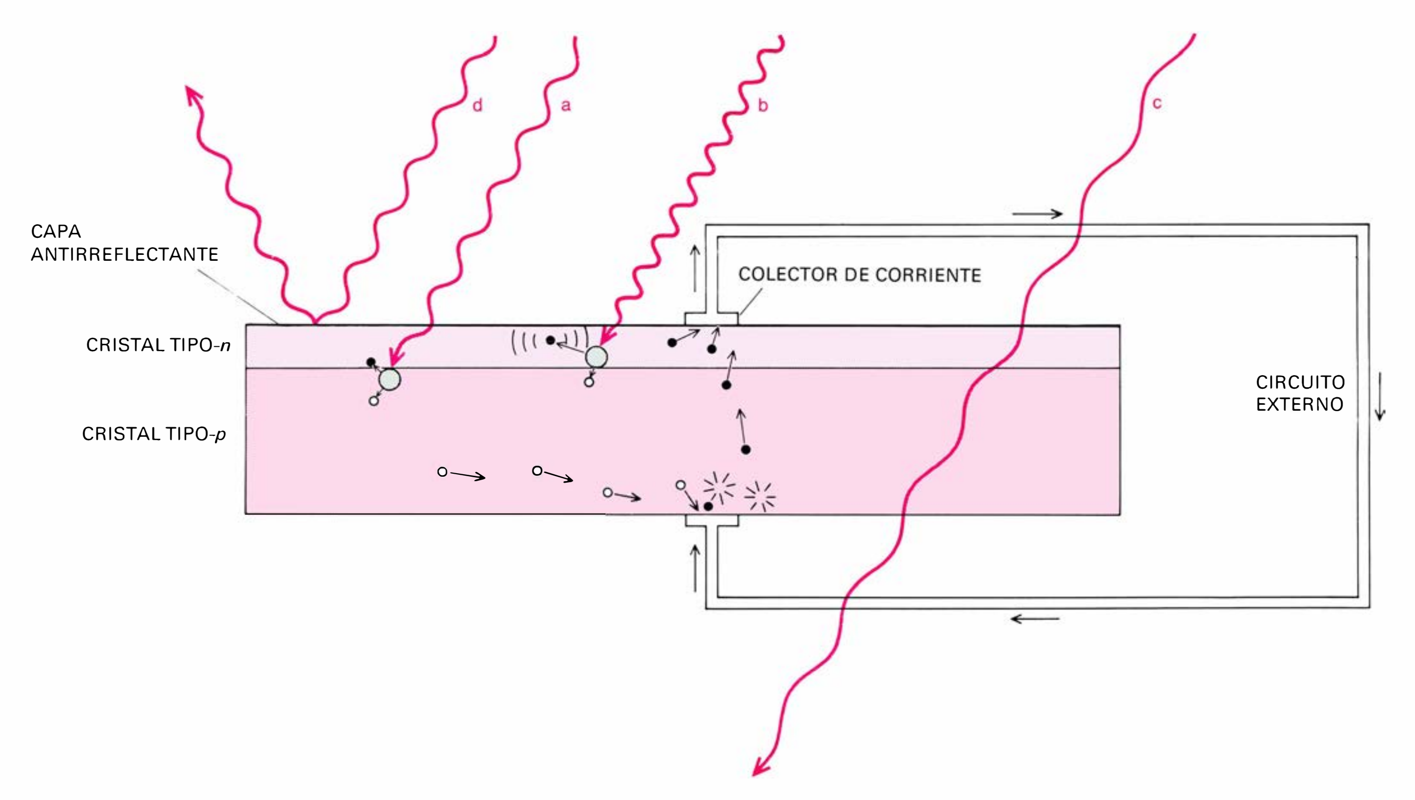
Se puede obtener una corriente eléctrica de la célula a través de un circuito que haga contacto eléctrico con las superficies anterior y posterior de la misma. A través de este circuito externo, los electrones atrapados en la zona tipo-*n* encuentran su camino de vuelta a la zona tipo-*p* donde se prodrán recombinar con los huecos. Si el circuito externo tiene una resistencia muy baja, la corriente que circula por él es una medida de la velocidad a la que los electrones se separan de los huecos. Esta depende de la intensidad de la luz incidente sobre la superficie de la célula y de la velocidad a la que electrones y huecos se pierden por recombinación.

¿Cuál es el rendimiento de un dispositivo fotovoltaico de silicio de este tipo? En condiciones óptimas, la energía solar llega a la superficie de la tierra a razón de unos 1,15 kilovatios por metro cuadrado. ¿Qué parte de esta energía puede convertirse en energía eléctrica? En términos más precisos, ¿qué rendimiento puede esperarse de un dispositivo fotovoltaico? De la misma forma que para determinar el rendimiento de una central térmica de fuel, hay que considerar tres puntos: 1) ¿Cuál sería el rendimiento del dispositivo si sus características fueran las ideales desde el punto de vista teórico? 2) ¿Hasta qué punto un dispositivo fotovoltaico que pueda realmente construirse se ajusta a las características ideales? 3) Un dispositivo fabricado a gran escala y precio aceptable, ¿cuánto se acerca al mejor dispositivo que podría fabricarse si no importara el precio?

Para considerar el rendimiento teórico de un dispositivo fotovoltaico, vamos a estudiar más de cerca el proceso por el cual un electrón absorbe un fotón en un semiconductor. Este proceso consiste en la transferencia de un cuanto de energía a un solo electrón. Si el fotón tiene energía suficiente, la energía recibida por el electrón liberará a éste de su normal función de ligadura entre dos átomos en el cristal; será libre de moverse en el cristal como un electrón de conducción y simultáneamente se creará un hueco.

La energía necesaria para hacer que un electrón pase de su estado ligado en el cristal a la banda de conducción se conoce como *gap* de energía. El gap de energía varía considerablemente de un semiconductor a otro; se puede expresar en electrón-voltios o en longitudes de onda de la luz que tiene la energía cuántica requerida. El silicio tiene un gap de energía de 1,12 electrón-voltios; por lo tanto, en un dispositivo fotovoltaico de silicio, un electrón necesita una energía de 1,12 electrón-voltios para pasar a la banda de conducción. Esta energía corresponde a una longitud de onda de 1,1 micras, que está en la región del infrarrojo del espectro.

Una radiación con una longitud de onda mayor que 1,1 micras no tiene energía bastante para hacer que los elec-



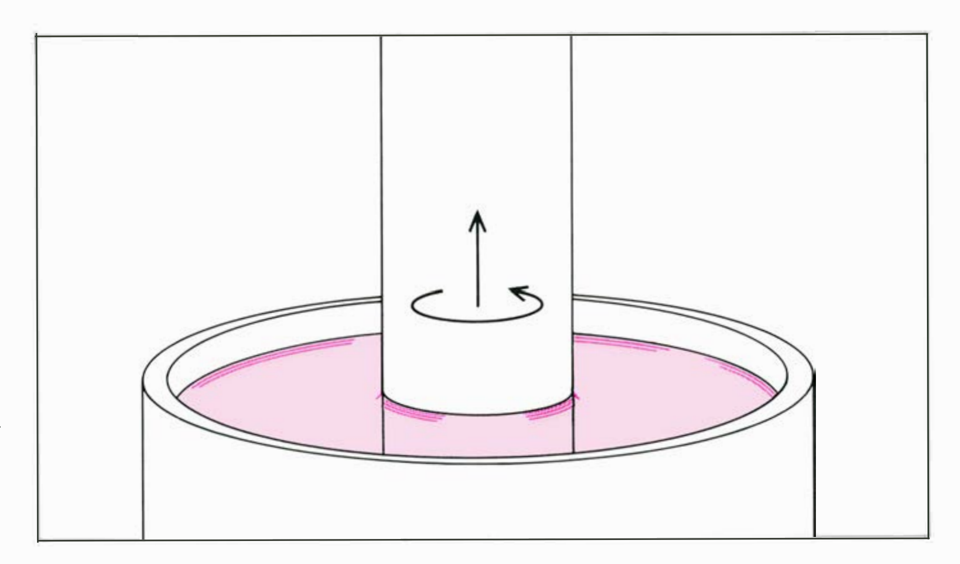
UNA CELULA SOLAR DE SILICIO es una pastilla de silicio tipo-*p* con una fina zona de silicio tipo-*n* sobre una cara. Cuando un fotón de luz con energía adecuada penetra en la célula cerca de la unión de los dos tipos de cristal y se encuentra con un átomo de silicio *(a)*; arranca uno de sus electrones, que deja tras de sí un hueco. La energía requerida para que un electrón pase a la banda de conducción se conoce por gap de energía. Este electrón excitado tiende a moverse hacia la zona de silicio tipo-*n* y el hueco hacia la zona de silicio tipo-*p*. Entonces, el electrón viaja hacia el colector de corriente que hay sobre la superficie frontal de la célula, genera una corriente eléctrica en el circuito externo y reaparece en la zona de silicio tipo-*p*, donde puede recombinarse con los huecos allí existentes. Si un fotón con energía superior a la del gap choca con un átomo de silicio *(b)*, también da lugar a un par electrón-hueco, y el exceso de energía se convierte en calor. Un fotón con energía menor que la del gap pasará directamente a través del cristal *(c)*, cediendo energía en su camino. Además, algunos electrones se reflejan en la superficie frontal de la célula aunque esta tenga recubrimiento antirreflectante *(d)*. Aún se pierden otros fotones porque no pueden alcanzar el cristal, ya que se lo impide el colector de corriente que cubre parte de la superficie frontal del cristal. Todas estas pérdidas suponen que una célula real de silicio no puede convertir en energía eléctrica más que, aproximadamente, un 19 por ciento de la energía solar que recibe, y parece poco probable que esta cifra pueda aumentar.

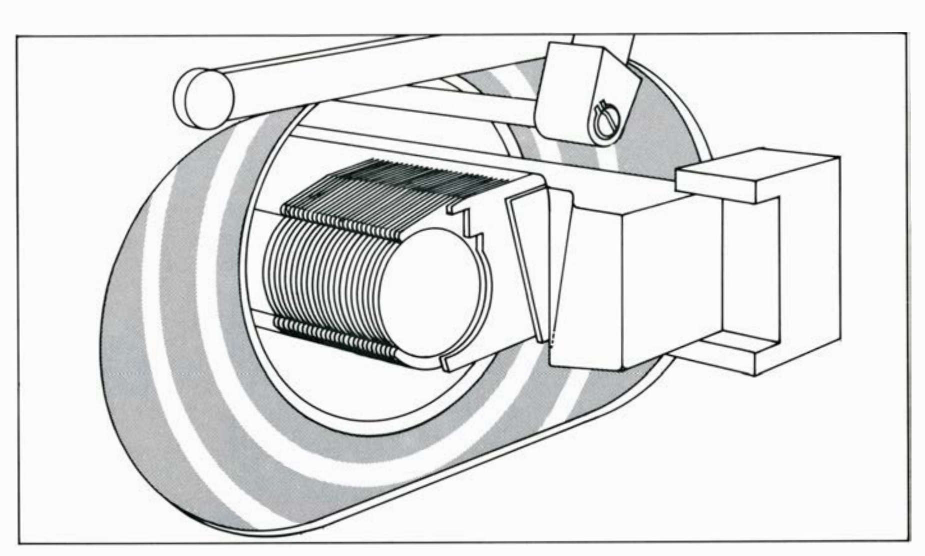
trones pasen a la banda de conducción en un dispositivo fotovoltaico de silicio. Casi la mitad de la energía que la tierra recibe del sol llega en forma de radiación con longitudes de onda mayores que 1,1 micras: por tanto, esta energía no interviene en la generación de electricidad en una célula de silicio. Si la energía del fotón es mayor que el gap de energía, esto es, si, como en el caso del silicio, la longitud de onda del fotón es menor que 1,1 micras, la energía queda absorbida completamente por un electrón. Sin embargo, el electrón no puede retener más de 1,12 electrón-voltios de la energía del fotón, y, en consecuencia cualquier exceso de energía se convierte inmediatamente en calor. En principio, si el fotón tuviese energía suficiente, podría hacer pasar dos electrones a la banda de conducción, pero de hecho todos estos fotones de alta energía se absorben por la atmósfera terrestre antes de alcanzar la superficie.

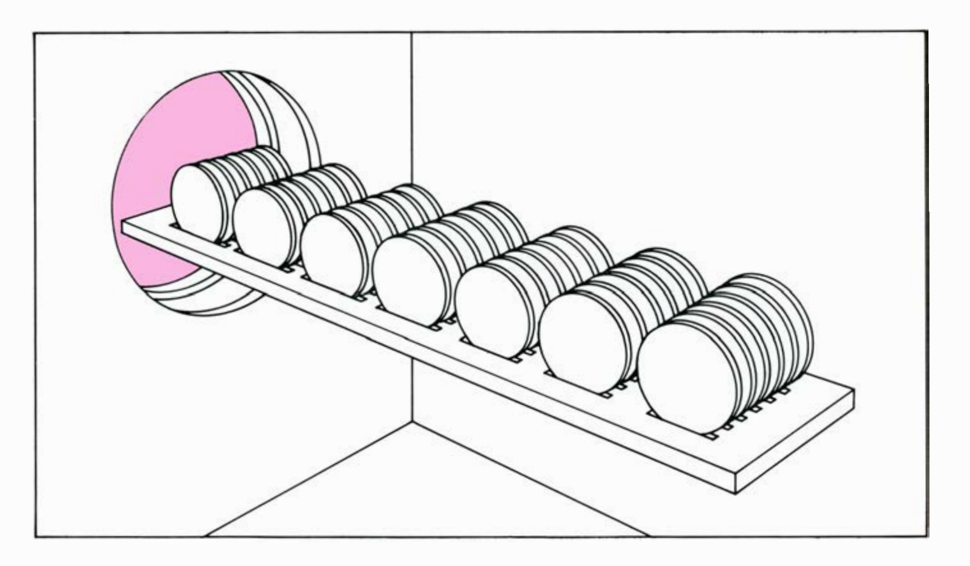
Como resultado de estas limitaciones físicas, sólo la mitad aproximadamente de los fotones que llegan a la tierra pueden crear pares electrón-hueco en una célula fotovoltaica de silicio y una parte considerable de la energía de estos fotones se convierte en calor en lugar de energía eléctrica. Para ser más precisos, las limitaciones fundamentales establecidas ya por la física cuántica restringen el rendimiento máximo posible de una célula solar de silicio en la tierra hasta un 21 %, aproximadamente. Otros semiconductores tienen diferentes gaps de energía y sus rendimientos teóricos máximos son, por tanto, diferentes.

Se puede dibujar una gráfica relacionando el rendimiento teórico y el gap de energía. La forma de la curva depende de la distribución de energía en el espectro de la radiación que se recibe: la curva de la página 23 es para luz solar fuera de la atmósfera terrestre que es la situación predominante en los vehículos espaciales. Sin embargo, cuando se trata de dispositivos en la superficie terrestre, debemos recordar que la atmósfera filtra la luz solar antes de que ésta alcance la superficie. Para cuando ha pasado a través de "la masa de aire 1", que sería su trayectoria vertical a través de la atmósfera, ha perdido una parte sustancial de su radiación de longitud de onda más corta (ultravioleta). La curva es, por tanto, algo diferente en sus detalles.

Aunque el máximo rendimiento teórico es una característica importante del semiconductor, hay muchas razones, menos fundamentales que la restricción cuántica mencionada, pero no menos







EL METODO CZOCHRALSKI para el crecimiento de cristales de silicio, consiste en hacer girar una semilla de cristal sumergida en un cristal de silicio fundido a la vez que se saca lentamente *(arriba)*. El resultado es un cristal cilíndrico de siete a diez cm de diámetro y hasta un metro de largo. Si inicialmente se añade a la fundición la cantidad adecuada de boro, el cristal queda uniformemente dopado. Para hacer células solares de silicio, éste se corta en pastillas finas *(centro)*; en este proceso, se pierde una gran cantidad de cristal en forma de polvillo. Una superficie de cada pastilla, que es de silicio tipo-*p*, se convierte en zona tipo-*n* al ser expuesta a fósforo *(abajo)* a temperatura suficientemente alta para que los átomos de fósforo se difundan hasta un pequeño hueco en la pastilla. Se hacen los contactos eléctricos, se aplica el recubrimiento antirreflectante y toda la célula se encierra en una cápsula protectora. El resultado es una célula fotovoltaica que opera con una eficacia entre el 15 y el 18 por ciento.

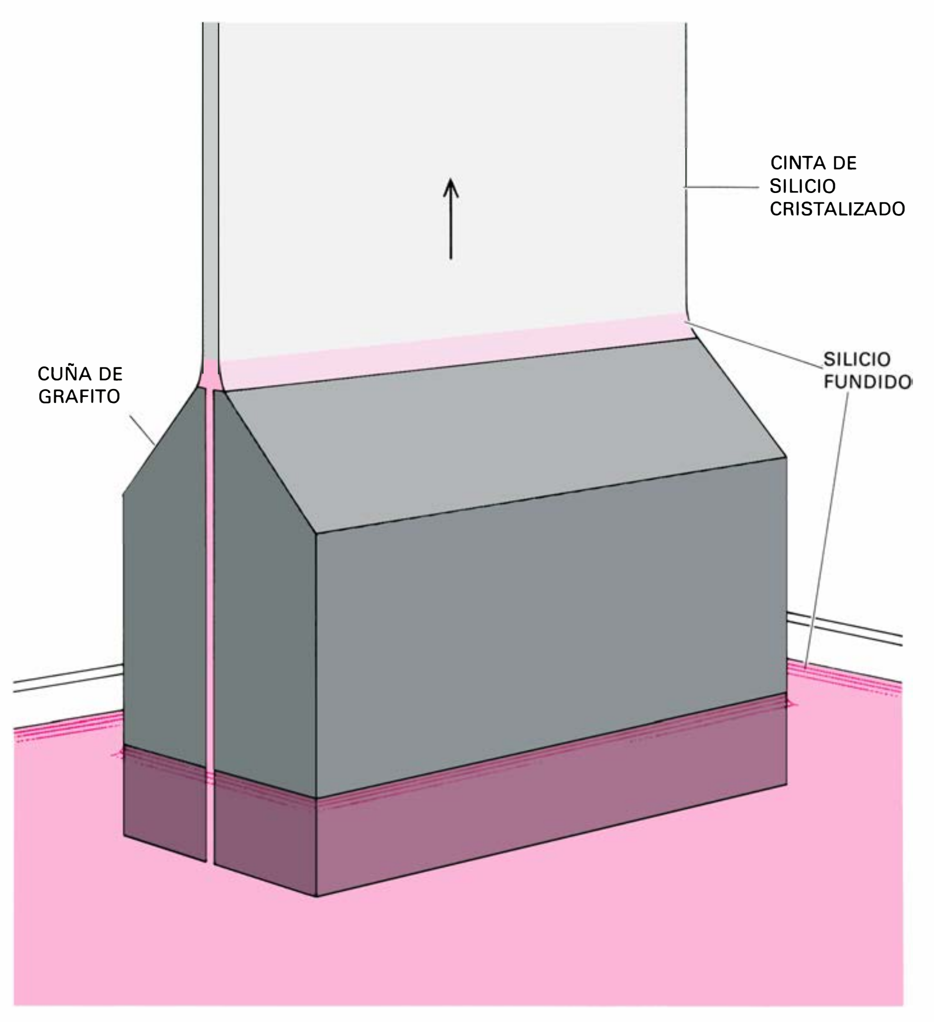
reales, que destruyen las esperanzas de alcanzar más de cuatro quintos, aproximadamente, de este rendimiento. Para que no se crea que un rendimiento de menos del 20 por ciento es algo indigno de nuestra técnica complicada, debería recordarse que el súmmum de la tecnología industrial avanzada, el automóvil, lleva a cabo su conversión de energía con un rendimiento menor que un 20 por ciento y realiza su función de transportar a la gente aun con menor rendimiento.

La consideración que lleva a un rendimiento teórico máximo del 21 por ciento para células solares de silicio no tiene en cuenta dos problemas que pueden ser minimizados pero nunca totalmente eliminados, y un tercer problema que puede reducirse en tanto en cuanto, sea económicamente rentable. El primer problema se basa en el hecho de que el 21 por ciento se refiere a la utilización de la energía de radiación que realmente ha entrado en el cristal, sin tener en cuenta que algo de la luz solar se refleja en la superficie del mismo. No hay forma de eliminar esta reflexión, pero puede minimizarse con un recubrimiento antireflectante: una capa de material transparente, de un espesor tal que los rayos de luz reflejados en la superficie del cristal interfieran destructivamente con los rayos reflejados en la superficie frontal del recubrimiento. Los rayos reflejados refuerzan entonces los rayos que penetran en el cristal. Sin embargo, incluso el mejor recubrimiento antirreflectante se puede diseñar de forma que se adapte sólo a una longitud de onda del espectro solar. Parece poco probable que la pérdida de energía solar por reflexión pueda llegar a reducirse por debajo del cinco por ciento.

El segundo problema es la resistencia del semiconductor a través de la cual la corriente debe pasar cuando va desde las proximidades de la unión entre las zonas tipo-*n* y tipo-*p*, donde se origina el voltaje, hasta los puntos en que la corriente entra en los conductores del circuito externo. El colector ideal de corriente consistiría en láminas de metal que cubrieran por completo las caras anterior y posterior de la célula solar, haciendo buen contacto con ellas. Dicho colector no presenta serios problemas para la parte posterior de la célula. Sin embargo, el conductor que cubriera la parte anterior de la célula tendría que ser transparente a la luz. Por razones fundamentales, los buenos conductores no son transparentes y, por consiguiente, no hay manera de cubrir la superficie frontal con una lámina que sea suficientemente conductora y al mismo tiempo permita que la luz alcance el cristal. Por el momento, deberá existir el compromiso de que el conductor recubra la menor parte posible de la superficie frontal de la célula, al mismo tiempo que se minimice la distancia entre cualquier punto en que se genere corriente y el punto más próximo donde se recoja.

Este compromiso presenta sus propios problemas. Si se intentara extraer un conductor partiendo de una red de metal extremadamente fino en contacto con la superficie del metal, sólo un pequeño porcentaje del área de la célula sería obscurecida. Sin embargo, la resistencia del colector sería muy alta, y se perdería energía debido al calentamiento de los hilos. Para reducir esta resistencia en una célula solar real, se cubre con el colector un diez por ciento, aproximadamente, de la superficie frontal; lo cual impide que un diez por ciento de la luz disponible alcance la célula. Por tanto, el rendimiento de la célula se ha distanciado en esta medida del rendimiento ideal. Hay aún otras pérdidas atribuibles a la resistencia que la corriente encuentra dentro del cristal cuando se dirige al colector, aunque la máxima distancia que debe atravesar no supere los tres milímetros.

Si una célula solar consistiera en un cristal perfecto de silicio, con las cantidades y distribuciones ideales de dopantes tipo-*n* y tipo-*p*, y sin otras impurezas, su rendimiento sería del 18 por ciento y ya se ha informado sobre las células con este rendimiento. Cualquier imperfección por impureza química o irregularidad de la estructura cristalina lleva a una disminución del rendimiento. Las impurezas pueden causar fugas de corriente a través de la unión en sentido inverso, y las imperfecciones cristalo-



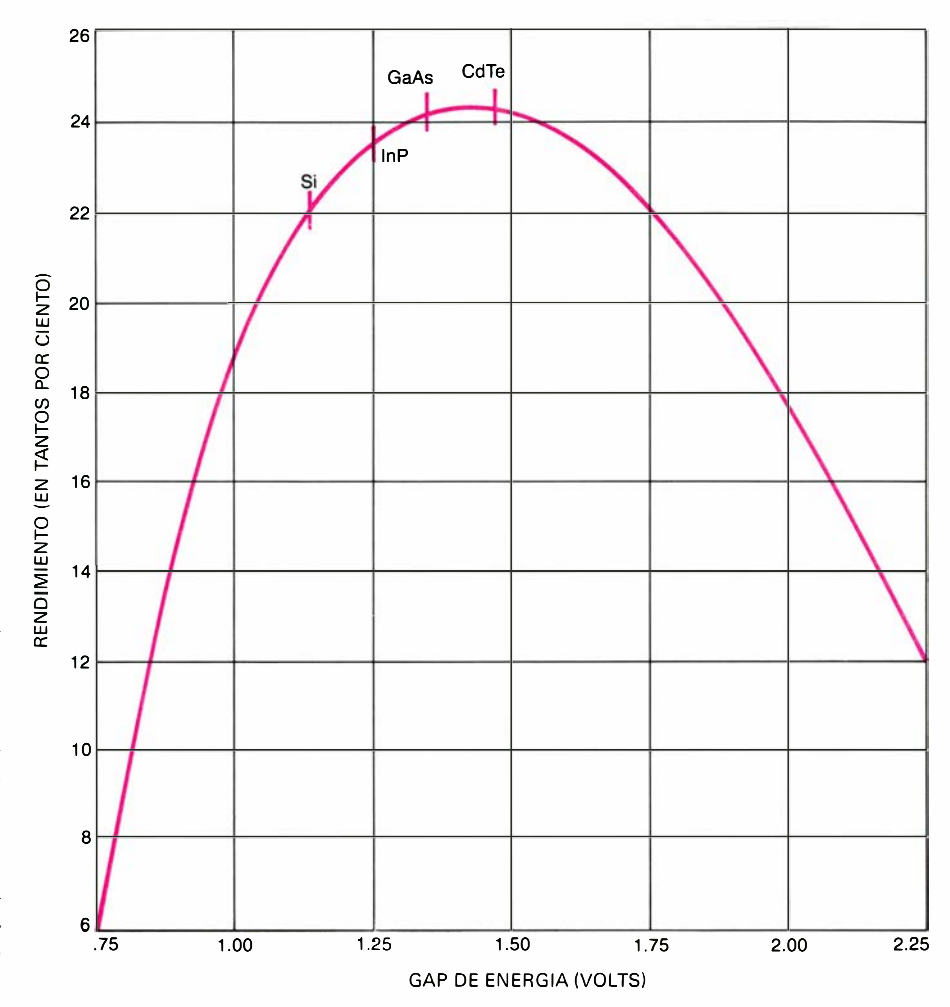
METODO EFG DE CRECIMIENTO de cristales de silicio mostrado en detalle. La cuña de grafito *(gris oscuro)* se mantiene en un baño de silicio fundido *(color)*. El silicio sube por capilaridad a través de una ranura estrecha de la cuña formando una lámina de silicio líquido en la parte superior de la cuña. Una semilla cristalina *(no indicada en la figura)* se sumerge en el líquido en la parte superior de la cuña, se saca luego a la misma velocidad que el cristal *(gris claro)* va desarrollándose hacia abajo, para que la interfase entre el cristal y el silicio líquido se quede algún centímetro por encima de la parte superior de la cuña. El cristal de silicio cristaliza en forma de cinta que puede cortarse a modo de pastilla con un mínimo de desperdicio.

gráficas pueden actuar como trampas donde los electrones y los huecos puedan recombinarse (y, por tanto, perderse) y como trayectorias de alta conductividad a través de las cuales pueden tener lugar cortocircuitos.

Las células solares descritas fueron desarrolladas para el programa espacial y proporcionaron la energía eléctrica a todos los vehículos espaciales. Constituyen una solución totalmente satisfactoria para el suministro de energía sin ningún tipo de combustible, pero son caras. Para comparar el coste de la potencia eléctrica generada por células solares con el coste de la potencia generada por las centrales eléctricas, el primer criterio es el capital invertido por kilovatio obtenido. Datos publicados en 1959 daban como capital invertido para las células solares 200.000 dólares por kilovatio. Actualmente, se pueden lograr para fines no espaciales por unos 20.000 dólares por kilovatio, pero la producción es muy limitada (llegando al equivalente de cien kilovatios al año). El coste de construcción de una central térmica de fuel es alrededor de 500 dólares por kilovatio de potencia eléctrica de salida. Obviamente el coste de las células solares es demasiado alto como para poder decir que no se necesitan combustibles.

¿Por qué es tan elevado el coste de los dispositivos fotovoltaicos? La razón estriba en la pureza química y la perfección cristalográfica requeridas para lograr que la célula solar alcance su más alto rendimiento. Los procesos normalmente empleados para lograr pureza química y perfección cristalográfica son caros en sí. Son los mismos procesos por los cuales se fabrican cristales para otros fines, tales como dispositivos microelectrónicos, en los que se requiere la más alta calidad posible. Si se trata de producir en gran escala células fotovoltaicas para convertir energía solar en energía eléctrica tendrán que conseguirse buenas células solares más baratas. A la vista de tal requisito, es razonable preguntarse: ¿es la fabricación de células solares intrínsecamente cara, o hay alguna posibilidad de reducir el coste en un factor de cuarenta, lo que las haría competitivas económicamente?

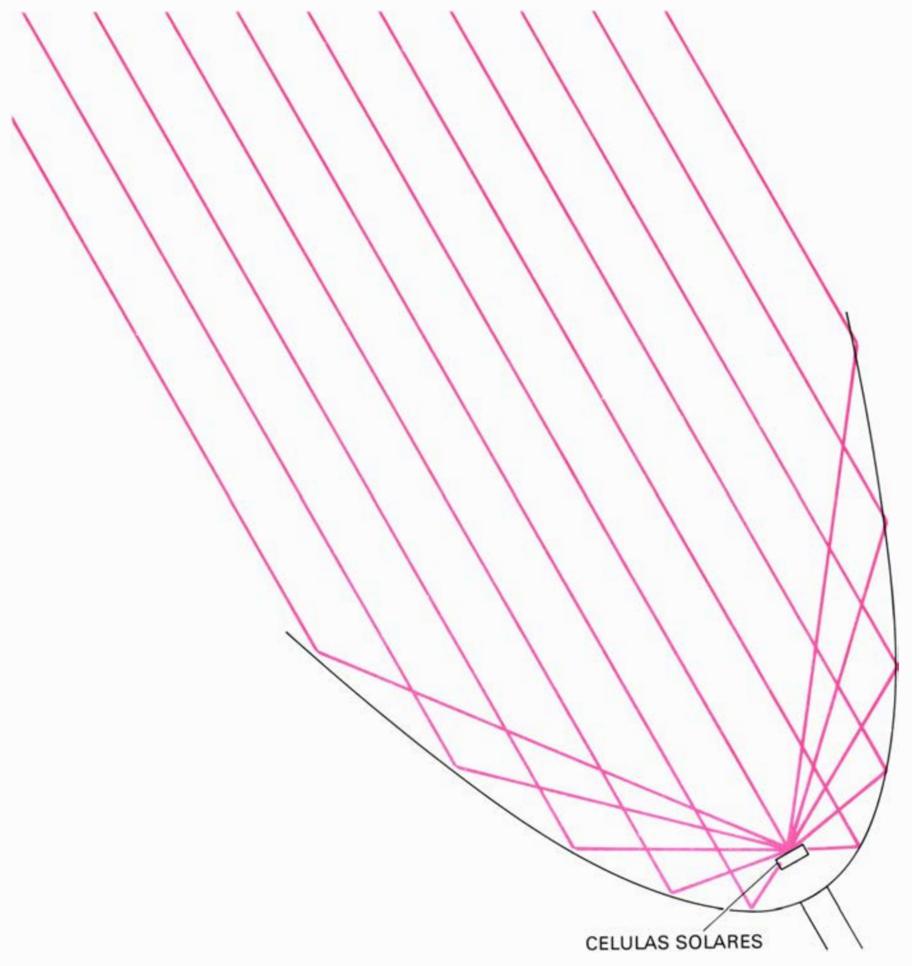
El proceso por el cual se fabrican células solares requiere, para empezar, silicio semiconductor, un material muy caro, cuyo coste es de 70 dólares por kilogramo, aproximadamente. Esto no se debe a la escasez del elemento (el silicio es uno de los elementos más abundantes en la corteza terrestre) sino a la complejidad del proceso de purificación. Cristales perfectos de silicio pueden desarrollarse por el método de Czochralski, un proceso por el cual una semilla de cristal gira al par que va saliendo lentamente de un crisol de silicio fundido. Si la temperatura y las velocidades de rotación y retirada se controlan con suficiente precisión, se obtiene un perfecto monocristal. Si se añade la cantidad de boro adecuada a la fundición, el cristal queda dopado en la medida requerida. El cristal así obtenido es bastante voluminoso: lo típico es un cilindro de siete a diez centímetros de diámetro y un metro de largo.



EL RENDIMIENTO TEORICO MAXIMO de una célula fotovoltaica depende del gap de energía del semiconductor de que está hecha. El gap de energía viene dado en electrón-voltios. Aquí, se representan los rendimientos de cuatro semiconductores: silicio (Si), fósforo de indio (InP), arseniuro de galio (Ga As) y teluro de cadmio (CdTe). Las restricciones de tipo fundamental de la Física limitan el rendimiento de la misma célula fotovoltaica a menos de un 25 por ciento.

El paso siguiente en la fabricación de las células solares reside en cortar el cristal en finas pastillas, cada una de las cuales será en su momento una célula. Para evitar que el cristal se dañe, las pastillas se sierran muy lentamente. Una pastilla típica posee un espesor de un cuarto de milímetro, y hasta el momento es un cristal uniforme de silicio tipo-*p*. Una fina capa de la misma se convierte en silicio tipo-*n* exponiendo una superficie de la pastilla a fósforo, a una temperatura suficientemente elevada como para que los átomos de fósforo se difundan hasta una profundidad de alrededor de media micra. Finalmente, se adhieren los contactos eléctricos a las caras superior e inferior de la pastilla, se aplica el recubrimiento antirreflectante y la célula completa se encierra en una cápsula protectora.

El producto de esta serie de operaciones es una célula que generará entre 10 y 15 milivatios por centímetro cuadrado. Puesto que el área de una célula típica es 45 centímetros cuadrados, si la célula operara con una eficacia típica de sólo un 15 por ciento, generaría dos tercios de vatio de electricidad expuesta a la luz solar. A precios normales, tal célula costaría alrededor de 13 dólares. Este alto costo es debido, en parte, a que una gran proporción del cristal inicial de silicio se desperdicia como polvo cuando se cortan las pastillas. Sin embargo, la



EL REFLECTOR PARABOLICO es un dispositivo destinado a concentrar la luz solar sobre una serie de células solares para disminuir el número de éstas. Tiene la desventaja de que, para su buen funcionamiento, necesita un mecanismo automático que lo mantenga apuntando al sol.

razón más importante del elevado precio de las células es que un proceso de fabricación tan preciso requiere tiempo y trabajo por parte de obreros especializados que desarrollan los cristales, cortan las pastillas y dan el acabado.

La potencia de salida de una célula solar de un tamaño dado es proporcional a la intensidad de radiación que incide sobre ella. Parece, por tanto, que sea económicamente rentable concentrar luz solar en un área relativamente pequeña de la célula. Esto puede conseguirse focalizando la luz solar por medio de un espejo parabólico. Sin embargo, este tipo de espejos tiene dos desventajas. En primer lugar, focalizaría la radiación de tal forma que incidiría en la célula desde varios ángulos, mientras que el mejor funcionamiento de la célula se da cuando la radiación incide formando ángulos rectos con la superficie frontal. En segundo lugar, sería necesario girar el reflector para mantenerlo apuntando al Sol a lo largo de todo el día, lo cual requeriría un sistema de guías automático bastante preciso.

Otra alternativa prometedora, en lugar del espejo parabólico, es el reflector de forma acanalada conocido como colector Winston. El colector está diseñado para reflejar una alta proporción de luz solar incidente sobre un área pequeña, incluso cuando el sol no está directamente sobre ellas y por tanto no es necesario que el colector siga al sol. Se ha demostrado que es factible una relación de concentración de ocho a uno, y que el funcionamiento de las células no sufre en tanto no se les permita calentarse demasiado. Se puede incrementar por tanto, la salida de un área dada de célula solar en un factor de ocho al menos, reduciendo el coste por kilovatio en el mismo factor. El que esto sea o no económicamente rentable depende del coste de los colectores Winston en relación con el de las células solares.

Pero, ¿es necesario mantener unos estándares de calidad tan extremadamente altos en los cristales de silicio para células solares? Es cierto que si un cristal está destinado a un dispositivo microelectrónico, la más leve imperfección puede hacerlo inútil. Sin embargo, los requisitos para una célula solar son menos rigurosos, y el deseo de alto rendimiento debe superarse con el elevado coste de la perfección. Como hemos visto, un cristal perfecto tendría un rendimiento de 18 por ciento, aproximadamente. Cristales de calidad algo más baja pueden tener rendimientos entre el 10 y 12 por ciento con un coste menor de la célula. El coste por kilovatio de salida sería entonces mucho más bajo. Se está haciendo un gran esfuerzo, financiado en parte por la Energy Research and Development Administration (ERDA), y en parte por compañías privadas, centrado en el desarrollo de métodos menos costosos de fabricación de silicio suficientemente puro y de conversión del mismo en pastillas destinadas a células solares. Uno de los métodos más prometedores consiste en el crecimiento continuo de una cinta de silicio de la anchura y espesor requeridos para las células solares. El proceso conocido como "crecimientos de láminas de bordes definidos" (edge-defined film-fed grouth) (EFG) nació en los Laboratorios Tyco, siendo desarrollado por Mobil Tyco Solar Energy Corporation. En este proceso, el silicio fundido sube por capilaridad, a través de una ranura, a una cuña de grafito. La película de silicio líquido sobre la parte superior de la cuña alimenta el crecimiento del cristal, cuya sección recta viene definida por la forma de la parte superior de la cuña. La cinta de silicio, de 0,15 cm de espesor, puede crecer a velocidades de unos 2,5 cm por minuto; entonces, simplemente se corta en pastillas de 2,5 cm de ancho y 10 cm de largo desperdiciando el mínimo de material. Los cristales crecidos por el método EFG son mucho menos perfectos que los crecidos por el método Czochralski, de ellos, sin embargo, se obtienen células solares con una eficacia de entre el 10 y el 12 por ciento.

El método EFG no es el único para producir cristales de silicio barato y a la vez suficientemente bueno para construir células solares, y, como he mencionado antes, no es el único material posible para tales células. Por ejemplo, se han fabricado células solares de arseniuro de galio de una eficacia del 18 por ciento. Además, láminas delgadas de otros semiconductores que se depositan por evaporación en un sustrato podrán algún día competir con las pastillas de silicio en términos de dólares por kilovatio. El método EFG aún no ha sido automatizado hasta el punto de que pueda producir células baratas. Sin embargo, un análisis profundo indica que tiene la posibilidad de producir cristales para células solares que costarían menos que la cifra mágica de 500 dólares por kilovatio.

Ahora bien, ¿cómo pueden emplearse exactamente tales células para convertir la intermitente y en cierto modo impredictible energía solar en energía eléctrica? En primer lugar, no hay que olvidar que el coste por kilovatio de las células no puede descender inmediatamente a niveles económicos, ya que los bajos costos no se logran hasta que existe producción a gran escala. Sin embargo, hay demandas en pequeña escala que podrían ayudar a la transición. El bombeo de agua para el riego o la alimentación de aparatos de televisión (con almacenaje de baterías) en partes del mundo lejanas y soleadas son aplicaciones para las cuales aun las células solares más caras pueden suponer la fuente de energía más barata. El mercado sería suficientemente extenso como para sostener el crecimiento inicial necesario para una nueva industria de células solares.

Sin embargo, esta etapa intermedia es relativamente insignificante en términos del programa nacional de energía de los Estados Unidos. Las cuestiones más importantes son: si los dispositivos fotovoltaicos pueden proporcionar energía suficiente para producir las importaciones de petróleo de Estados Unidos y el número de centrales nucleares que se construyen, y si pueden, cuándo sucederá.

Hay tres posibilidades muy distintas. La primera es que se podrían construir grandes centrales de energía solar en los desiertos del suroeste de Estados Unidos donde el sol es abundante. Con un rendimiento del 10 por ciento un área colectora de 10 km² (un cuadrado de 3,16 km de lado) podría generar una potencia máxima de mil megavatios, la salida típica de una gran central térmica. Aunque la salida de las células solares es en forma de corriente continua podría convertirse en corriente alterna y alimentar el circuito de distribución de una empresa eléctrica.

Sin embargo, la electricidad suministrada de esta forma no podría reemplazar otros medios de generación de electricidad ya que no podría utilizarse en todo momento en que se necesitase Las centrales de energía solar podrían, no obstante, reducir la demanda de centrales térmicas y, por tanto, ahorrar combustible, cubriendo parte de la demanda mientras el sol brilla. Tal función se llama, a veces, almacenamiento negativo.

Por otra parte, las centrales de energía solar podrían generar electricidad cuando el sol brillara y almacenarla para su uso cuando no hubiera sol. La energía podría almacenarse bombeando agua a depósitos elevados; la energía se recuperaría al permitir caer el agua a su nivel inicial a través de turbogeneradores. Parece, también, posible almacenar energía solar fabricando hidrógeno mediante la electrolisis del agua. Se almacenaría el hidrógeno y posteriormente se quemaría en lugar de gas natural, o podría servir de combustible para producir electricidad.

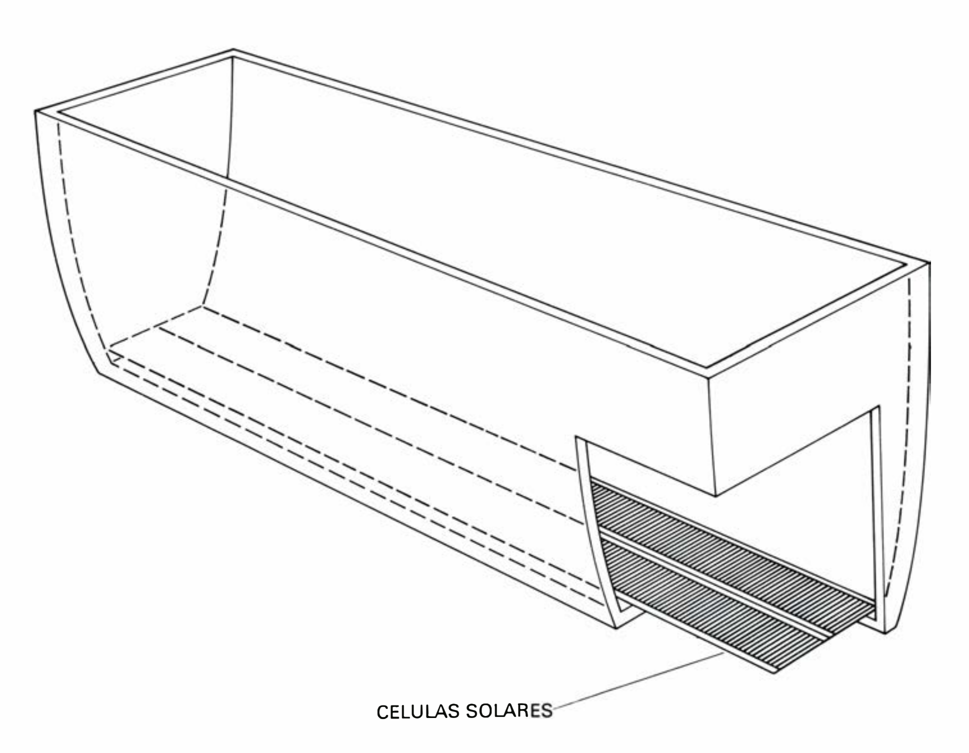
Otra forma por la cual los sistemas fotovoltaicos podrían sustituir a las fuentes de energía más convencionales podría ser el despliegue de grandes paneles de células solares en estaciones espaciales en órbitas fijas alrededor de la Tierra. La energía recogida del sol se transmitiría a las estaciones receptoras terrestres por medio de haces de microondas. Este sistema tendría la ventaja de proporcionar energía continuamente (si se utilizase más de una estación espacial) y no habría necesidad alguna de almacenamiento.

Además, las células recogerían la energía solar en el espacio, donde es dos veces más intensa que en la superficie terrestre. No hay duda que tal estación espacial podría construirse para la obtención de energía pero requeriría una enorme inversión. También podría objetarse que las microondas podrían ser peligrosas para la vida en la tierra y que el sistema sería vulnerable en manos de un aprovechado enemigo.

La industria de aparatos eléctricos se ha desarrollado en una era en que la economía de generación de energía requiere una producción de energía altamente centralizada. Este requisito resulta del hecho de que es totalmente ineficaz generar energía en cantidades del orden de 1 kilovatio quemando combustible.

La forma más económica de generar electricidad con los métodos modernos es con centrales de energía que operan a niveles del orden de mil megavatios. Sin embargo, tales consideraciones no sirven para las células solares. Para ellas, la economía de escala se aplica a la producción de las células pero no a su utilización, ya que tiene el mismo rendimiento sea grande o pequeña la cantidad de electricidad.

Se puede argumentar que sería más barato que la electricidad necesaria en una casa fuera generada por células solares colocadas en el tejado y almacenada en baterías. Las células solares podrían integrarse en un sistema que proporcionaría calefacción, aire acondicionado y agua caliente, utilizando parte del 88 al 90 por ciento de la energía solar que las células solares no pueden convertir en electricidad. Además, si la electricidad fuese generada en el mismo lugar que se necesita, se eliminaría el cos-



EL COLECTOR WINSTON, inventado por Roland Winston, de la Universidad de Chicago, semeja un largo canal. Concentra la luz solar sobre una cinta de células solares y tiene la ventaja de que no necesita girarse para seguir al sol a lo largo del día, dada la condición de su diseño.

te por distribución al consumidor. No es probable que tal sistema fotovoltaico pudiese suministrar toda la energía necesaria para una casa. Sin embargo, si el sistema suministrara dos tercios de esta energía a un coste razonable, la demanda de centrales de energía se reduciría y la comunidad sería menos dependiente del suministro de petróleo, gas natural, carbón y uranio.

Y ¿cuándo podremos empezar a recoger los beneficios de esta tecnología? Depende en parte de la rapidez con que los dispositivos que funcionan en los laboratorios puedan fabricarse a escala industrial, lo cual depende de la rapidez con que se alcance el desarrollo tecnológico necesario. No se necesitan nuevos descubrimientos científicos, sólo una gran dosis de meticulosa ingeniería para poner en práctica ideas ya probadas. Aunque hoy en dia dispusiéramos de un proceso completo en todos sus detalles para fabricar células solares al coste de 500 dólares por kilovatio de salida, harían falta cinco años para montar una industria que produjera las células solares y el equipo asociado para generar electricidad a escala de megavatios, y podrían pasar otros cinco años antes de que una parte significante de la generación de energía fuera por medios fotovoltaicos.

El objetivo de la ERDA es conseguir, para los primeros años de la década de los 80, varios sistemas fotovoltaicos experimentales de diez megavatios, así como una contribución significativa, quizás superior al 1 por ciento, a nuestro consumo de energía (equivalente a casi un millón de barriles de petróleo por día), para finales de siglo. La capacidad de generación fotovoltaica podría acelerar su desarrollo, pero requeriría mucho más capital y mucho más esfuerzo tecnológico de los que probablemente se les ha de destinar. Sin embargo, la demostración de que es factible la generación de electricidad a gran escala con células solares no garantiza su amplia adopción; esto dependerá de quien pueda vender el producto de su sistema de generación eléctrica a un precio relativamente más bajo.

El coste de 500 dólares por kilovatio para sistemas fotovoltaicos se considera como un propósito realista, pero ¿es válido este criterio para considerar dichos sistemas económicamente competitivos? En principio, los 500 dólares por kilovatio incitan al interés, pero debe recordarse que una central de energía convencional (costando también 500 dólares por kilovatio) puede generar electricidad y, por consiguiente, dar beneficios al capital invertido, de día y de noche, llueva o haga sol. Las estadísticas oficiales ponen de manifiesto que, en promedio, las centrales de energía de los Estados Unidos funcionan el 50 por ciento del tiempo aproximadamente. Esto significa que se generan alrededor de 4400 kilovatios-hora al año por kilovatio instalado. Por otra parte, para un generador solar, la electricidad generada por kilovatio de salida sería aproximadamente 2100 kilovatios-hora al año, en Texas, y 1400 kilovatios-hora al año en los estados del noreste. Por tanto, si una instalación de energía solar tuviese que proporcionar la misma energía que una central de petróleo de 1000 megavatios o que una central nuclear, necesitaría una salida máxima entre 2000 y 3000 megavatios.

El precio que tendrá en el futuro el combustible determinará si un factor de dos o tres en el coste quedará compensado por el coste nulo del combustible en una instalación de energía solar.

Los factores económicos se encuentran más favorables, con mucho, a los generadores solares en aquellos lugares en los que la demanda de electricidad es máxima, porque los acondicionadores de aire funcionan a tope. En dichos lugares es donde, al mismo tiempo, la energía solar es más abundante. Sin embargo, la habilitación de generadores solares como equipo adecuado para el servicio bajo estas condiciones nunca podría suponer más que una pequeña contribución a la demanda total de energía eléctrica. A la hora de emplear electricidad procedente de generación solar, habrá que tener en cuenta otras consideraciones de tipo económico.

Dado que una fracción notable de lo que cuesta al consumidor la electricidad viene representado por su distribución, a un coste de 500 dólares por kilovatio de salida, el coste por kilovatio-hora que hubiera de pagar el abonado estaría muy por debajo del de las compañías eléctricas.

Actualmente, una central fotovoltaica que costase 500 dólares por kilovatio de salida (si pudiera construirse) no sería competitiva con ninguna central alimentada con petróleo o uranio. Sin embargo, al aumentar el precio y las restricciones de dichos combustibles y, por tanto, el coste de generación de electricidad a partir de ellos, la energía solar se hará más atrayente. La generación fotovoltaica de electricidad puede encontrar un lugar bajo el sol antes de lo que muchos sospechan.

Estudios sobre los ribosomas mediante dispersión de neutrones

Los orgánulos de la célula viviente donde se forman las proteínas constan de 58 macromoléculas. Irradiando un ribosoma con un haz de neutrones, se determina la distribución espacial de sus moléculas

Donald M. Engelman y Peter B. Moore

De dos maneras se puede llegar a la comprensión de la estructura biológica. Un método, ya establecido hace tiempo, consiste en analizar los materiales vivientes a niveles cada vez más finos, examinando primero los tejidos, luego las células y finalmente las partes componentes de las mismas. El principal instrumento para estas investigaciones ha sido el microscopio, tanto el óptico como el electrónico, éste más reciente. El segundo método sigue la dirección opuesta, a saber: desde las estructuras más pequeñas hasta las mayores, acotando la disposición de los átomos en las distintas moléculas de una complejidad creciente. En este empeño, la técnica más importante es el estudio de figuras de difracción por rayos X; durante los últimos 20 años este tipo de estudios ha ido revelando la disposición espacial de las átomos en unas cuantas docenas de proteínas, lo cual constituye un logro realmente notable.

Entre estos dos dominios, queda un espacio en blanco: algunos componentes de la célula son demasiado pequeños para poder determinarse sus estructuras con la microscopía electrónica, pero demasiado grandes para ser reconstruidos a nivel atómico a partir de estudios de difracción por rayos X. Un ejemplo de ello es el ribosoma, un orgánulo encontrado en todas las células (comúnmente en gran número) en el cual el código genético es traducido a su expresión final; que es la secuencia de aminoácidos en una molécula proteica. El tamaño de los ribosomas los hace asequibles a la micrografía electrónica, pero sólo lo suficiente para apreciarse su silueta; la información estructural que puede obtenerse de esta micrografía resulta, pues, limitada. Por otra parte, el ribosoma es mucho mayor que las estructuras que hasta el presente se han encontrado adecuadas para su análisis mediante difracción por rayos X. El ribosoma consta de 55 moléculas distintas de proteínas (cada una de las cuales forma una compleja entidad por sí misma) y de tres moléculas de ARN, todas de gran tamaño.

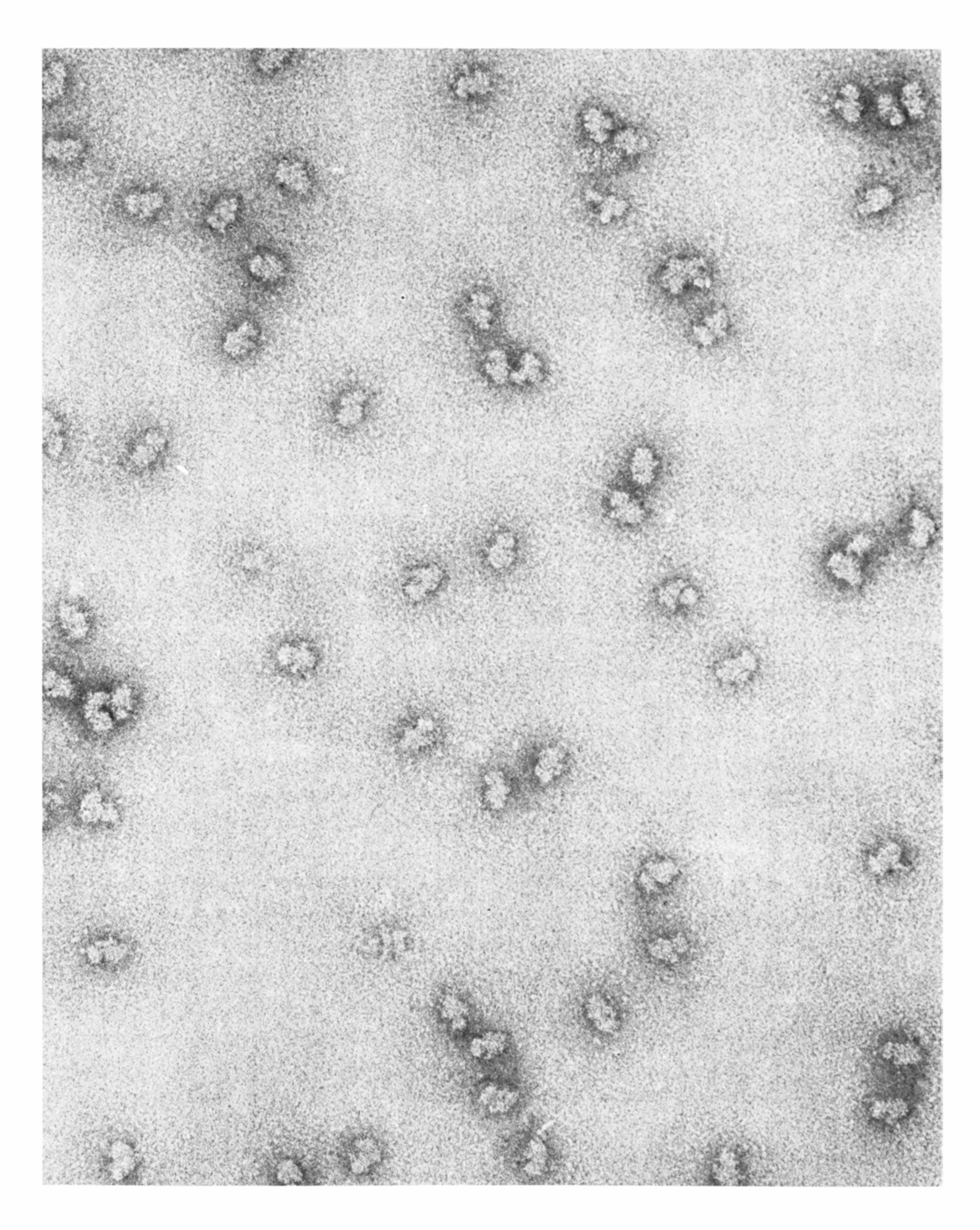
La información estructural en esta escala intermedia es esencial para comprender cómo están organizadas en unidades funcionales las macromoléculas biológicas del tipo de las proteínas. En el caso del ribosoma, difícilmente puede comprenderse su función determinante en la síntesis proteica, sin conocer previamente la naturaleza de la agregación entre sus partes componentes.

Durante los cuatro últimos años, hemos venido desarrollando un método para determinar la estructura de los complejos macromoleculares de ese tamaño; y hace un año, aproximadamente, comenzamos la aplicación del método al estudio de una de las subunidades de la estructura del ribosoma. La base de nuestro método es la dispersión de un haz de neutrones por subunidades ribosómicas especialmente preparadas. El resultado final de la investigación, si se ve colmada por el éxito, será un mapa en donde aparezca la disposición en el espacio de todas las proteínas de la subunidad.

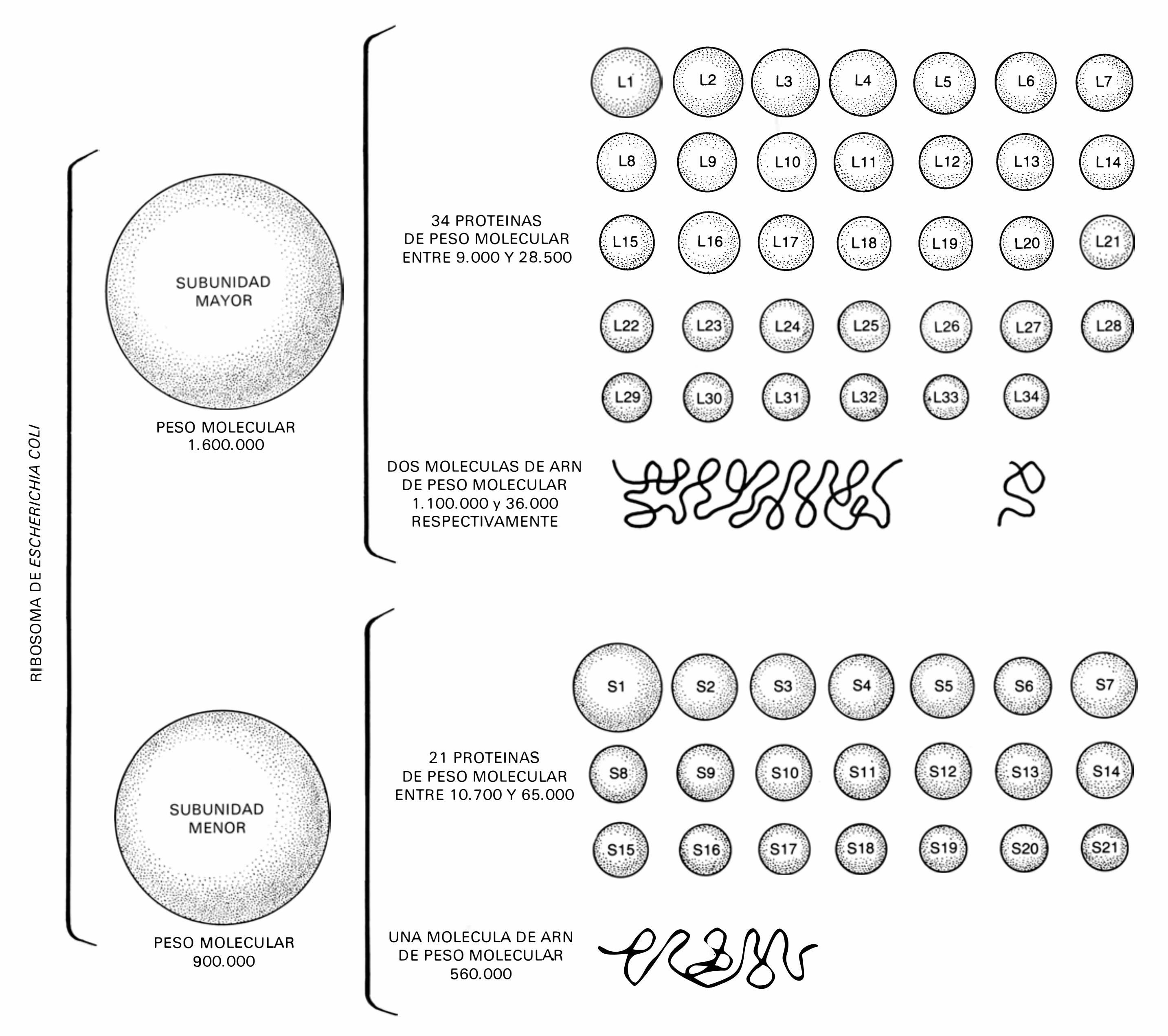
El "dogma central" de la biología molecular establece que la información en la célula fluye desde el ADN hacia el ARN y las proteínas. El código genético, representado por la secuencia de bases en una molécula de ADN, primero se copia para producir una molécula de ARN, y éste pasa del núcleo al citoplasma en forma de ARN mensajero. El ARN mensajero sirve entonces como modelo para la síntesis de una proteína, síntesis que sólo puede llevarse a cabo en presencia de los ribosomas. Los ribosomas coordinan la traducción del código: interaccionan con el ARN mensajero y reúnen las otras moléculas necesarias para la ordenada elongación de la cadena proteica. Entre tales moléculas, destacan como principales las unidades componentes de la proteína que va formándose, a saber: los aminoácidos; cada aminoácido es portado por una especie determinada de ARN de transferencia, que es otra clase de ARN. Se requiere también la presencia de varios factores que cumplen una función catalítica; la energía consumida en la formación de enlaces entre los aminoácidos es suministrada por el enlace fosfato, rico en energía, que ofrece una molécula pequeña llamada guanosín trifosfato.

El ribosoma ocupa obviamente una posición central en el flujo de información que va desde los genes hasta las proteínas, pero sólo tenemos una idea rudimentaria de cómo funciona. Se sabe que consta de dos subunidades, llamadas mayor y menor, que se enlazan por separado con el ARN mensajero al comienzo de la síntesis proteica. La subunidad mayor consta de 34 moléculas proteicas distintas y de dos moléculas de ARN ribosómico (una mayor que la otra) y tiene un peso molecular total de 1,6 millones aproximadamente. La subunidad menor se compone de 21 proteínas y de una molécula de ARN; su peso molecular es algo inferior al millón. Esta subunidad menor constituye el objeto de nuestros estudios.

En principio, las subunidades ribosó-



LAS PARTICULAS RIBOSOMICAS son los cuerpos irregulares que muestra esta microfotografía electrónica de transmisión, de J. A. Lake, de la Facultad de Medicina de la Universidad de Nueva York. Cada partícula es un componente, llamado subunidad menor de un ribosoma de la *Escherichia coli.* El análisis de las microfotografías electrónicas puede dar una idea de la forma de la subunidad, pero estas imágenes ofrecen poca información sobre su estructura interna. Por otra parte, las subunidades son de un tamaño excesivo como para ser investigadas convenientemente por ciertos métodos, como el del estudio de figuras de difracción por rayos X, que dan un mapa de su estructura átomo por átomo. Los estudios por dispersión de neutrones dan información variable de rasgos estructurales del componente ribosómico *E. coli.*



LAS MOLECULAS QUE COMPONEN el ribosoma de *E. coli* constan de 55 proteínas y de 3 moléculas de ARN. Estas moléculas están organizadas formando dos subunidades; de momento, sólo la subunidad menor ha sido analizada por dispersión de neutrones. La subunidad menor consta de 21 proteínas, con pesos moleculares que oscilan entre 10.700 y 65.000 y de una única molécula de ARN, con un peso molecular de 560.000. El análisis determina cómo están ordenadas estas moléculas en el espacio, considerando sólo las proteínas. Las proteínas y las subunidades se configuran como esferas cuyos volúmenes se relacionan con sus pesos moleculares, pero muchas no son esféricas.

micas, o el ribosoma entero, podrían estudiarse mediante difracción por rayos X. Tales estudios requieren, sin embargo, que el espécimen esté en forma cristalina; y de momento no ha sido posible conseguir cristales de ribosomas o de sus subunidades. Incluso en el caso de que pudiera obtenerse el cristal, el análisis de la figura de difracción por rayos X, generada por una estructura tan compleja, presentaría enormes dificultades. Los componentes proteicos de las subunidades, que teóricamente deberían cristalizar con mayor facilidad, quedan como candidatos para el análisis estructural mediante rayos X.

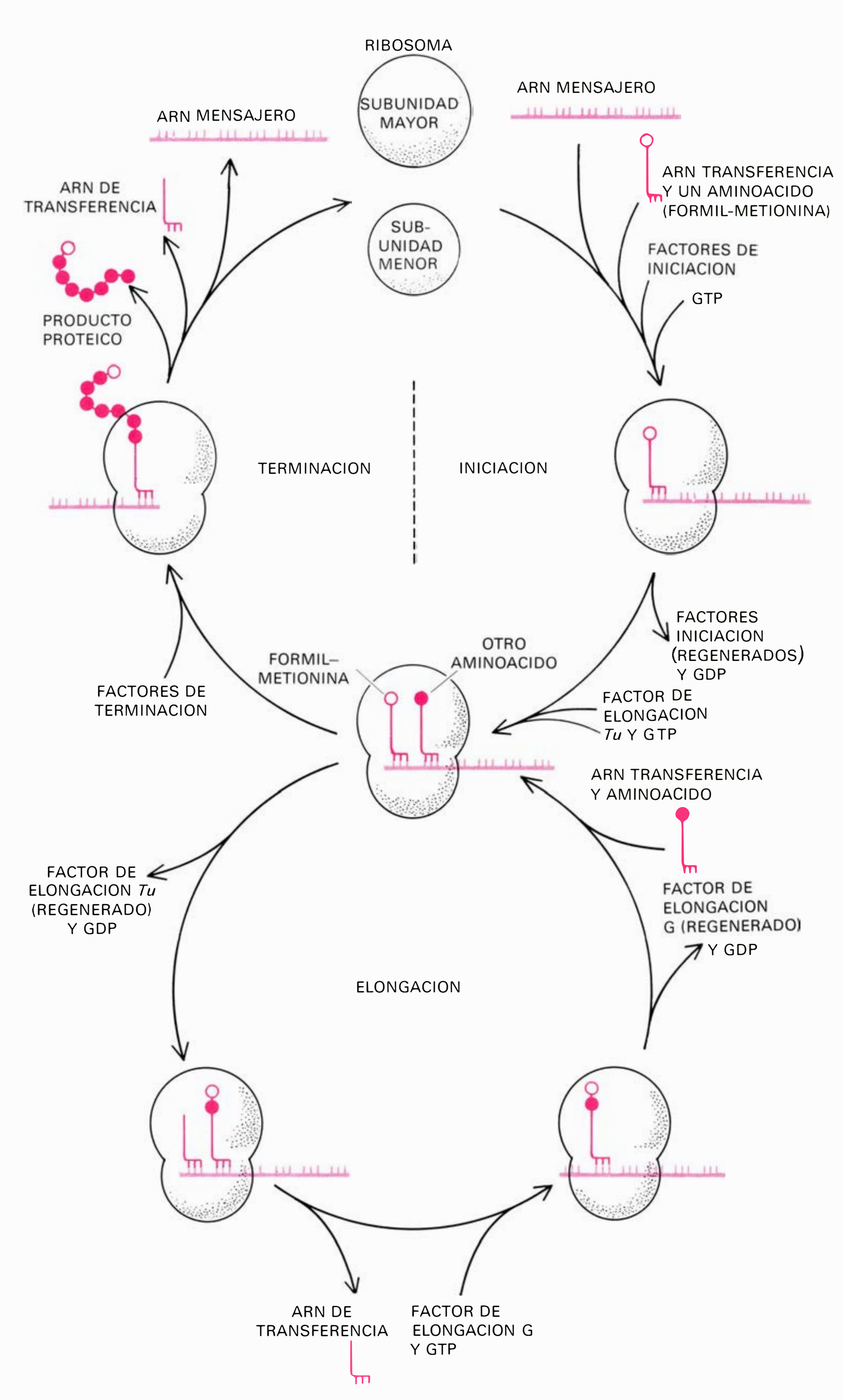
En vez de recurrir a la cristalografía por rayos X, se han desarrollado varios métodos, aparte del nuestro, para abordar el análisis estructural. El primero en dar un resultado positivo fue el empleo de agentes químicos dotados de dos grupos reactivos que se unían a moléculas adyacentes en la estructura. Rompiendo los ribosomas y determinando qué moléculas se habían enlazado, se podía compilar por lo menos una secuencia parcial de las proteínas en razón de su vecindad. Este sistema se ha venido empleando en varios laboratorios, concretamente en el laboratorio de Robert R. Traut, en la Universidad de Davis, California, y en el de Charles G. Kurland, en la Universidad de Uppsala, Suecia. Georg Stöffler, del Instituto Max-Planck de Genética molecular de Berlín y James A. Lake, de la Facultad de Medicina de la Universidad de Nueva York, propusieron una técnica inmunológica. Esta se fundaba en que los anticuerpos específicos para ciertas proteínas ribosómicas se unen a los ribosomas, formando unos complejos que pueden examinarse con el microscopio electrónico. Al poder determinarse la posición del anticuerpo, quedaba al descubierto así la posición que la correspondiente proteína ocupa en la estructura.

En otra línea de investigación, llevada a cabo por Charles R. Cantor y sus colegas en la Universidad de Columbia, se marcaron dos proteínas seleccionadas con moléculas de colorantes fluorescentes, sensibles a diferentes longitudes de onda. Se estimulaba entonces a uno de los colorantes iluminándolo con una radiación de longitud de onda apropiada; la eficacia con que su energía de excitación se transfería al otro colorante medía la distancia entre las dos proteínas. Otros métodos se desarrollaron registrando la dispersión de rayos X bajo condiciones que no requerían un espécimen cristalizado. Todos estos métodos pueden contribuir al conocimiento de la estructura de los ribosomas.

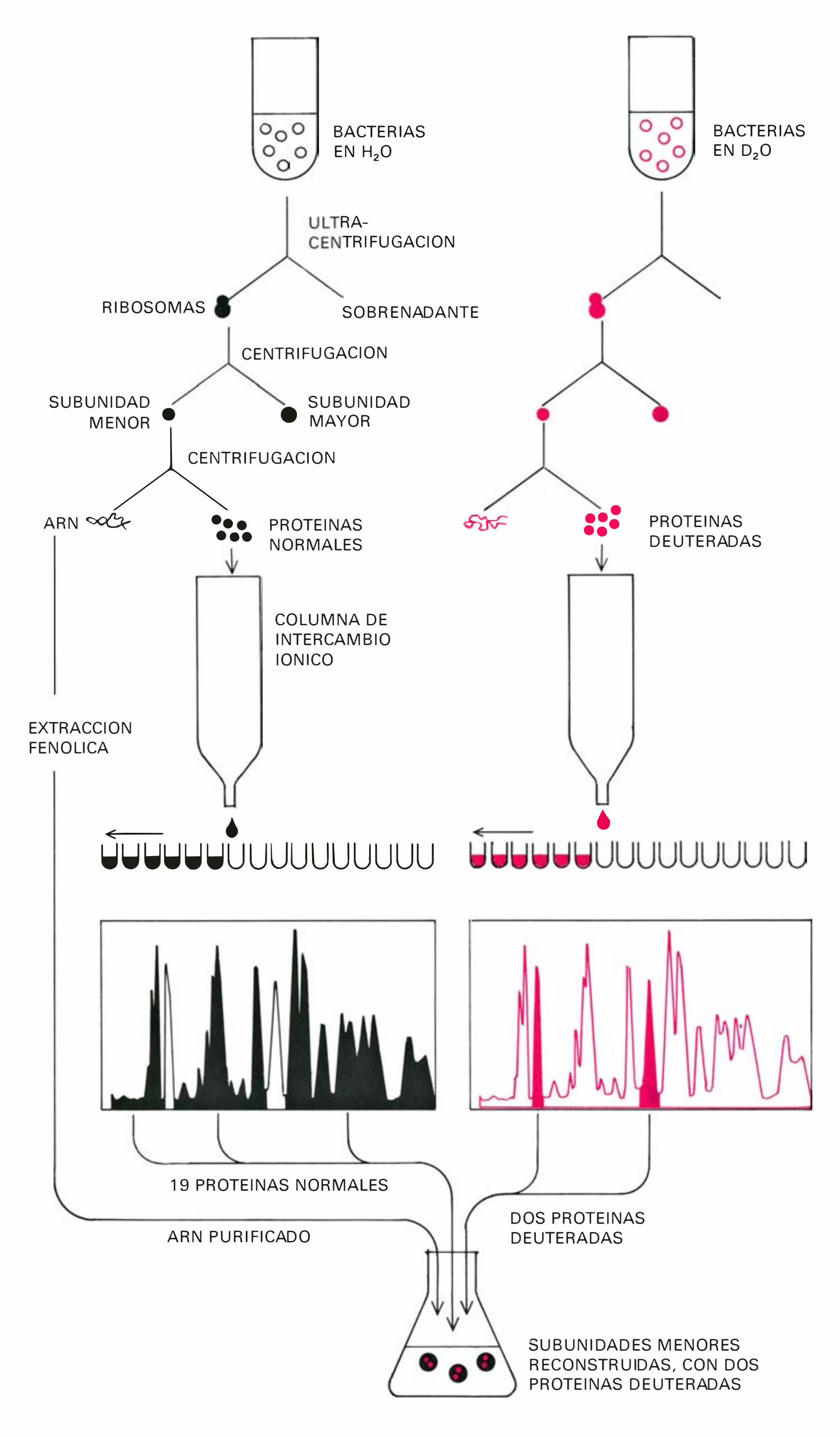
Nuestro método es puramente físico; no depende directamente de las propiedades biológicas o químicas del ribosoma. Los experimentos físicos suelen ser de diseño sencillo, aunque, como se verá, no siempre son fáciles de llevar a cabo. En nuestro experimento se interpone una solución de subunidades ribosómicas en un haz de neutrones. Los neutrones interaccionan con la muestra y, como resultado, cambia la dirección de propagación de algunos de ellos. Estos cambios de dirección informan de la estructura de la muestra; el nudo de la cuestión estriba en detectar los cambios y en interpretar su información.

Prácticamente todas las investigaciones sobre la estructura biológica han empleado luz visible, electrones o rayos X. Para el caso de la luz y de los haces de electrones se pueden construir lentes y, por ende, formar una imagen de la estructura sometida a inspección. Los rayos X, por el contrario, no pueden ser enfocados; no se puede construir un microscopio de rayos X. En lugar de una imagen, los rayos X producen sólo una figura de difracción, un registro de las direcciones e intensidades de las ondas desviadas al interaccionar con el espécimen. Al igual que los rayos X, los neutrones tampoco pueden ser enfocados y únicamente pueden registrarse sus ángulos de dispersión y sus intensidades.

Los neutrones difieren de todos los demás tipos de radiación empleados en las investigaciones biológicas por un hecho importante. La luz, los electrones y los rayos X interaccionan casi exclusivamente con los electrones en una molécula. Los neutrones no quedan alterados en presencia de electrones, sino que interaccionan con el núcleo atómico. Por consiguiente, los neutrones pueden ser dispersados de distinta manera por átomos que representan dife-



LA FUNCION BIOLOGICA del ribosoma es intervenir en la traducción del código genético para formar la secuencia de aminoácidos en una proteína. Él está representado por la secuencia de nucleótidos en una molécula de ARN mensajero. En la indicación, el ribosoma reconoce el punto de partida en un segmento del ARNm y enlaza con una molécula de ARN de transferencia portadora de un aminoácido. En todas las proteínas bacterianas, este primer aminoácido es la formil-metionina. En la elongación se enlaza un segundo aminoácido al primero. El ribosoma desplaza luego su posición sobre la molécula del ARNm y se repite el ciclo de elongación. Al alcanzar el final del mensaje genético, la cadena de aminoácidos se pliega espontáneamente para formar una proteína. Después, el ribosoma se divide en dos subunidades, que se reunirán antes de que se traduzca un nuevo segmento de ARN mensajero. La síntesis proteica viene facilitada por ciertas proteínas catalíticas (los factores de iniciación, elongación y terminación) y por el guanosín trifosfato (GTP), una pequeña molécula que libera energía al convertirse en guanosín difosfato.



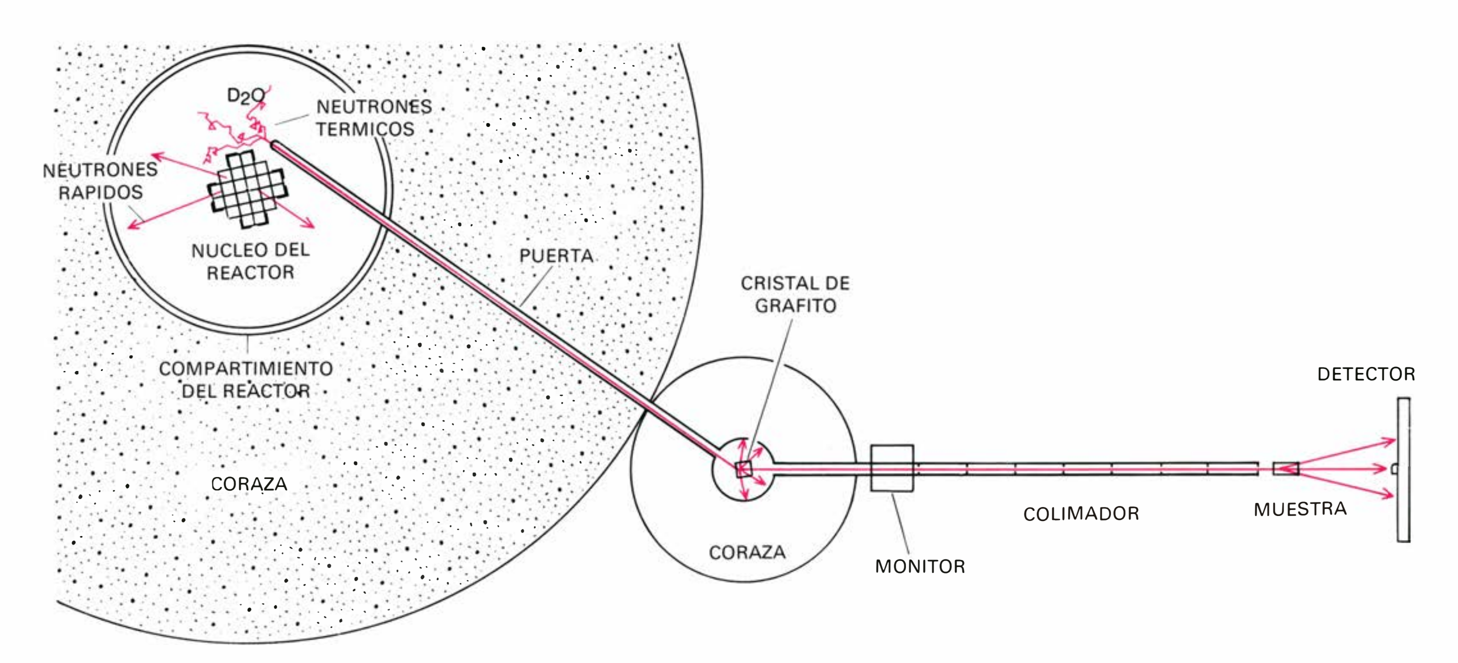
LA PREPARACION DE LAS SUBUNIDADES ribosómicas menores, para su análisis estructural mediante dispersión de neutrones, requiere "teñir" dos de las 21 proteínas componentes con deuterio, el isótopo del hidrógeno con un peso atómico de 2. Las bacterias se cultivan tanto en agua normal como en óxido de deuterio, o agua pesada (D_2O). En las proteínas derivadas de las bacterias crecidas en agua pesada, prácticamente todas las posiciones que le correspondería ocupar al hidrógeno, están ocupadas por el deuterio. Se extraen los ribosomas de las bacterias, se separan las subunidades menores de los ribosomas y se trae el ARN ribosómico de las subunidades menores, quedando un conjunto de 21 proteínas. Estas proteínas, tanto las correspondientes a los ribosomas normales como a los deuterados, se separan luego y se purifican mediante una técnica llamada cromatografía de intercambio iónico. Del conjunto de proteínas normales se descartan dos; seleccionándose las dos correspondientes del conjunto de las proteínas deuteradas. Las 19 proteínas normales y las dos deuteradas se mezclan junto con el ARN ribosómico normal; y las subunidades menores se reconstruyen espontáneamente. Las subunidades contienen dos proteínas marcadas con deuterio, y éstas difieren de las normales por su capacidad para dispersar los neutrones.

rentes isótopos del mismo elemento. Tales átomos tienen el mismo número de electrones, y, por lo tanto, un haz de electrones o una radiación electromagnética apenas los pueden distinguir; en sus núcleos tienen el mismo número de protones, pero distinto número de neutrones. Entre los isótopos de un elemento dado, hay sólo diferencias inapreciables en las propiedades químicas, pero puede haber grandes diferencias en la interacción de sus núcleos con neutrones de energía adecuada; se dice que los isótopos tienen diferentes capacidades para dispersar neutrones.

Una de las mayores diferencias en la capacidad para dispersar neutrones reside entre el hidrógeno y el deuterio, dos isótopos cuyos núcleos contienen un protón solitario y un protón y un neutrón, respectivamente. Como las moléculas biológicas contienen un gran número de átomos de hidrógeno, hay una gran diferencia en el comportamiento de la dispersión neutrónica entre una proteína que contiene sólo hidrógeno y otra en la que la mayor parte de los átomos de hidrógeno ha sido sustituida por átomos de deuterio. Sin embargo, dos moléculas de este tipo difieren muy poco en cuanto a sus propiedades químicas. El deuterio, por lo tanto, se convierte en el marcador adecuado para los experimentos biológicos de dispersión de neutrones. Reemplazando todos los átomos de hidrógeno en una molécula proteica por átomos de deuterio, la proteína se hace conspicua al haz de neutrones, a pesar de que sus propiedades químicas apenas se hayan alterado.

En la práctica experimental, se preparan las subunidades ribosómicas menores de manera que dos de las 21 proteínas estén fuertemente deuteradas. Las subunidades, en solución, se colocan entonces bajo el haz de neutrones, contándose luego el número de neutrones desviados según distintos ángulos. La información más significativa que se desprende de esta operación es simplemente la medición de la distancia entre las dos proteínas deuteradas.

Por los años veinte, Louis de Broglie sugirió que la materia y en particular los átomos y sus constituyentes podían considerarse como partículas o como ondas. La elección de una u otra interpretación es, frecuentemente, pura cuestión de conveniencia. La naturaleza de nuestro experimento postula, por razones de mayor utilidad, la consideración de los neutrones como ondas y no como partículas. Pues los fenómenos que estamos observando son los de difracción e interferencia de los neutrones, y estos

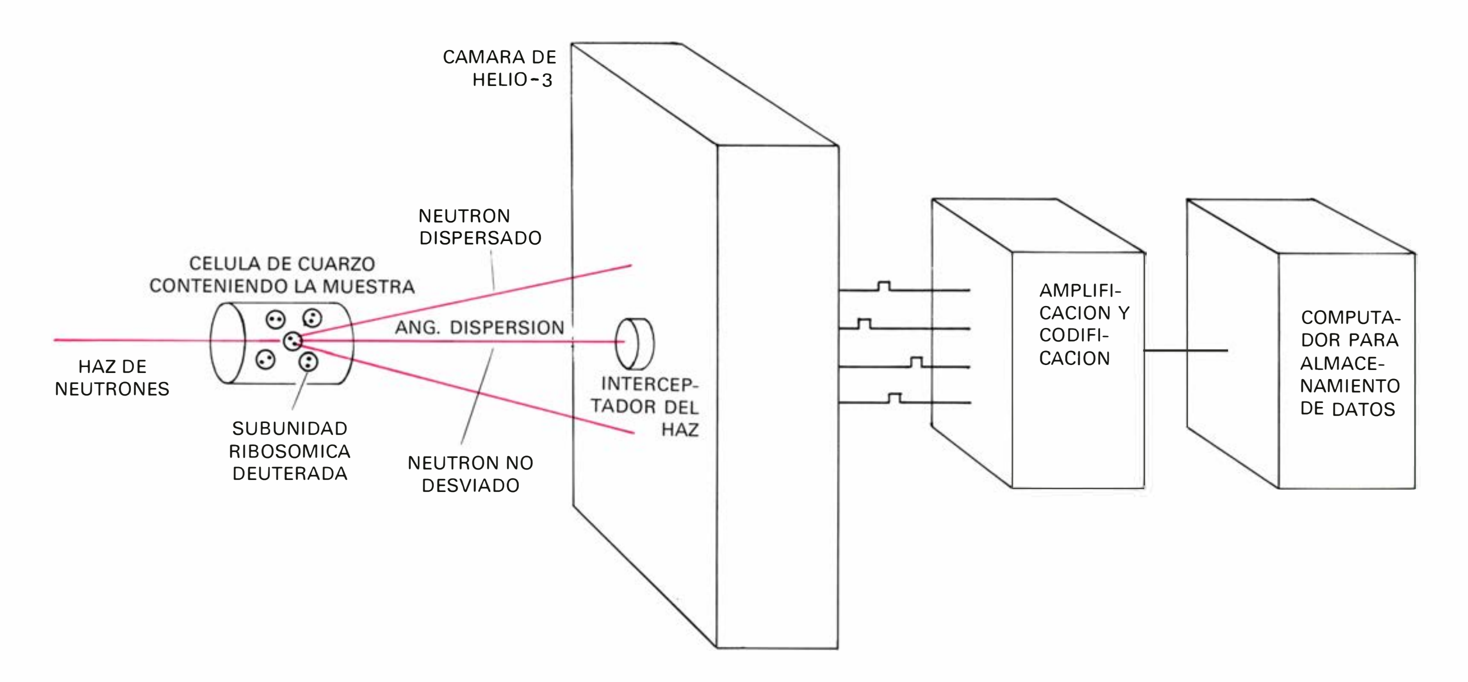


EL HAZ DE NEUTRONES empleado en los estudios sobre el ribosoma se genera en el High Flux Bearn Reactor del Brookhaven National Laboratory. Los neutrones son emitidos por la fisión de núcleos de uranio 235 en el núcleo del reactor. Luego, se ven frenados por las colisiones con los núcleos del manto de agua pesada que rodea el núcleo, y salen del reactor por una puerta del compartimiento. La distribución de energías entre estos neutrones retardados es como la de un gas en equilibrio térmico; se denominan neutrones térmicos. El haz de éstos es dispersado según sus energías al hacer reflexión sobre los planos atómicos de un cristal de grafito; se selecciona un rayo en el cual todos los neutrones tienen la misma energía mediante un colimador. El rayo colimado choca con el espécimen y los neutrones con el detector.

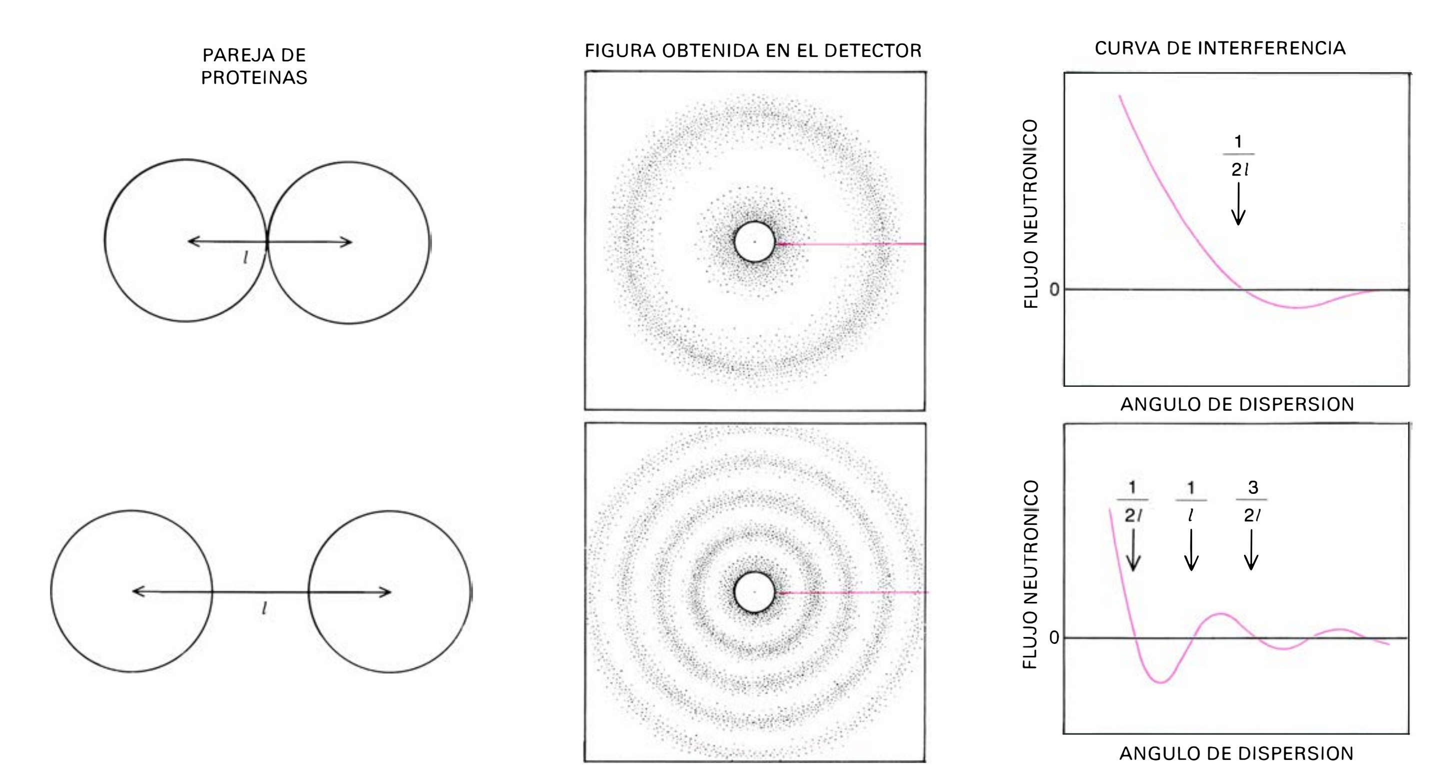
efectos están generalmente asociados a las radiaciones ondulantes.

Los neutrones empleados en nuestro experimento tienen una longitud de onda de algunos angstroms, determinada por su energía y, consiguientemente, por su velocidad. Cuando estas ondas neutrónicas pasan cerca de las proteínas deuteradas, se difractan o desvían; además, como cada subunidad ribosómica contiene dos proteínas deuteradas, las ondas difractadas por cada par se interfieren posteriormente.

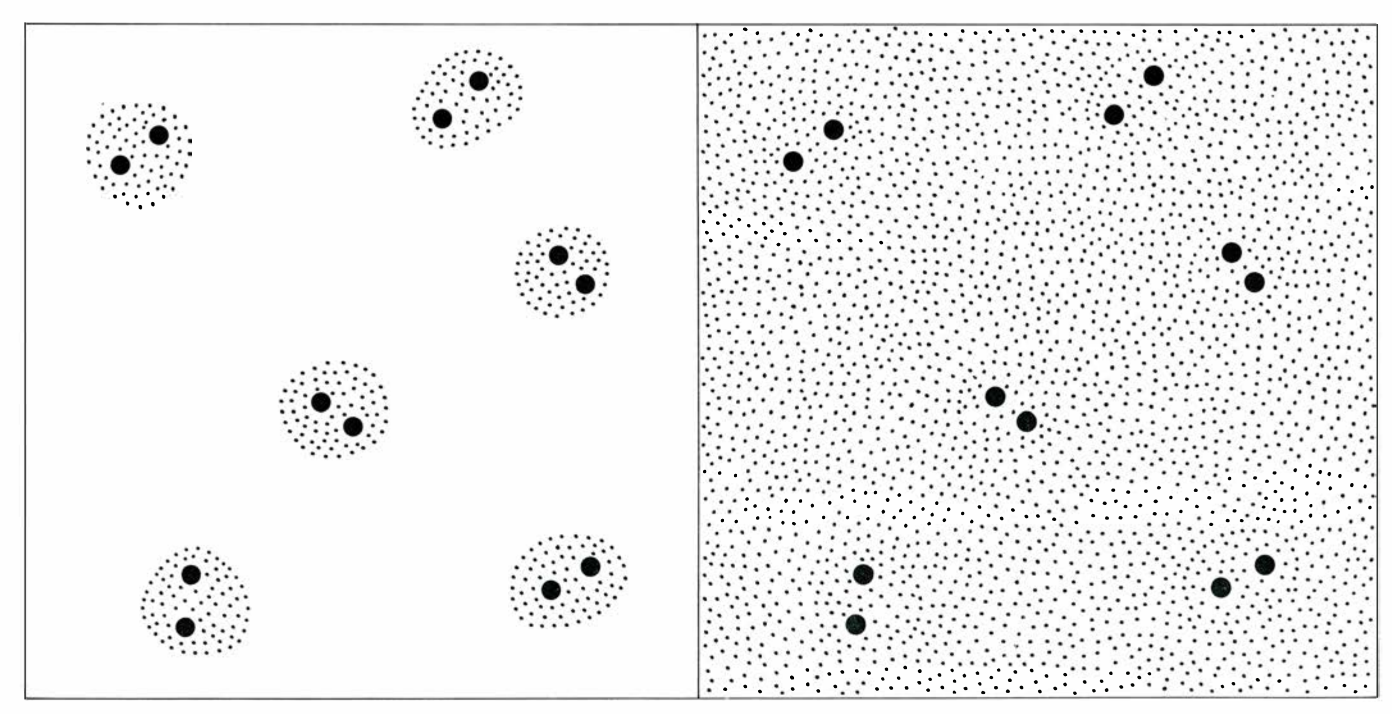
Cuando las ondas luminosas emitidas por una única fuente de luz pasan a través de un par de hendeduras paralelas practicadas en una pantalla opaca, las ondas originan una figura de interferencia que consta de varias franjas o bandas paralelas, claras y oscuras alternativamente. El efecto observado en nuestros experimentos de dispersión de neutrones es análogo. Las dos proteínas



EL DETECTOR DE NEUTRONES registra la intensidad del flujo neutrónico en cada una de más de 8000 posiciones sobre su superficie; de esta manera mide la distribución angular de los neutrones dispersados por las subunidades ribosómicas deuteradas. El detector es una cámara plana llena de helio 3, un isótopo del helio que emite un protón cada vez que recibe el impacto de un neutrón. La carga eléctrica del protón rechazado induce una ionización momentánea en su vecindad inmediata siendo registrada por un sistema de electrodos distribuidos en el plano del detector. Las señales detectadas por los electrodos son amplificadas, codificadas con el fin de denotar la posición del neutrón incidente y registradas en cinta magnética para su procesado. El rayo de neutrones no desviados es detenido por el interceptor del haz.



LA FIGURA DE INTERFERENCIA es generada por los neutrones dispersados a su paso por las subunidades ribosómicas que contienen dos proteínas deuteradas. El fundamento de los estudios del ribosoma mediante la dispersión de neutrones es el análisis de dichas figuras con el fin de determinar la distancia entre los centros de un par de proteínas. La figura de interferencia se presenta como un conjunto de anillos concéntricos, en los cuales el flujo de neutrones es alternativamente más alto y más bajo que el nivel de fondo; puede representarse por una gráfica del flujo de neutrones en función de la desviación angular del rayo. En el caso más simple, en que las dos proteínas se consideran esferas perfectas, la gráfica es una curva sinusoidal amortiguada. La distancia entre sucesivos anillos de interferencia está en relación inversa con la distancia entre las proteínas. Si se consideran las proteínas como esferas en contacto, los anillos están muy separados *(arriba)*; al aumentar la separación entre las proteínas, los anillos de interferencia se acercan más unos a otros *(abajo)*. La distancia se suele tomar desde el centro de la figura hasta el borde del primer anillo, esto es, donde la curva de interferencia corta el cero por primera vez. La ondulación de la curva de interferencia tiene una amplitud prácticamente pequeña, viene a ser un pequeño tanto por ciento del flujo total de neutrones.



LA IGUALACION DE CONTRASTES es una técnica que viene empleándose para mejorar la señal de interferencia en el proceso de dispersión de neutrones. La interferencia resulta del contraste efectivo (para los neutrones térmicos) entre las proteínas deuteradas *(puntos gruesos)* y las proteínas normales *(puntos más finos)* en cada subunidad ribosómica. Sin embargo, si las subunidades están en suspensión en agua pura, el haz de neutrones también detecta el contraste entre la subunidad como un todo y el medio, mediante la aparición de un componente espúreo en la señal de interferencia. Esta fuente de ruido puede eliminarse ajustando las propiedades de dispersión neutrónica del medio, de tal manera, que se igualen a las propiedades de las subunidades. En consecuencia, a la izquierda, los objetos más conspicuos son los agregados en forma de puntos, cada uno de los cuales representa una subunidad ribosómica; mientras que a la derecha son más aparentes los pares de puntos más gruesos, cada uno de los cuales representa un par de proteínas deuteradas, desaparecidas ya las propias subunidades.

deuteradas equivalen a las dos hendeduras, y la figura generada consiste en anillos concéntricos en los cuales el flujo de neutrones es, alternativamente, mayor y menor que el del nivel de fondo. En el experimento óptico, la distancia entres dos bandas, para cualquier longitud de onda luminosa dada, viene determinada por la distancia entre las hendeduras; en el experimento con dispersión de neutrones, la distancia entre los anillos está en relación con la distancia existente entre las proteínas deuteradas.

Sin embargo, hay ciertas diferencias entre nuestro experimento y el análogo óptico. En el nuestro no hay un par de hendeduras, sino muchos pares, que corresponden al gran número de ribosomas que hay en la muestra; estos pares están orientados al azar con respecto al haz de neutrones, pues los ribosomas están en solución. La teoría en que se basa la interpretación de una figura de difracción de este tipo fue elaborada en 1915 por Peter J. W. Debye. Debye estaba interesado en describir la dispersión de rayos X por un gas diatómico; pero la dispersión de neutrones en nuestro experimento es bastante similar. En ambos casos, las ondas son difractadas por objetos que están orientados al azar, que tienen dos centros de dispersión separados por una distancia fija.

La teoría de Debye establece que una gráfica de la intensidad de la radiación dispersada con respecto a la distancia al eje del rayo encierra un sector de una curva sinusoidal oscilante, que se amortigua para ángulos de desviación grandes. O sea, se observarán bandas de interferencia que serán prominentes para ángulos pequeños y débiles para ángulos grandes. Lo más importante para nuestro propósito es que la distancia entre los anillos de interferencia están en relación inversa a la distancia entre los centros de dispersión. Si los centros de dispersión –en nuestro caso las proteínas deuteradas– están juntos, la curva de interferencia presenta oscilaciones lentas, es decir, picos y depresiones muy espaciados; si los centros de dispersión están más separados, la curva de interferencia oscila más rápidamente o, en otras palabras, las franjas están más juntas.

No es tarea fácil preparar subunidades ribosómicas menores en las cuales dos proteínas seleccionadas contengan una alta concentración de deuterio. No sería posible en absoluto si no mediara una extraordinaria propiedad de la misma subunidad. En 1968, Peter Traub y Masayasu Nomura, de la Universidad de Wisconsin, descubrieron que bajo ciertas condiciones específicas, la subunidad menor se reorganiza espontáneamente a partir de sus partes componentes. Por esta razón, si pudiéramos preparar y aislar las moléculas componentes necesarias, habría bastantes posibilidades de construir la subunidad completa, según nuestras especificaciones. Decidimos trabajar con la subunidad menor porque se conocía un procedimiento para su reconstrucción.

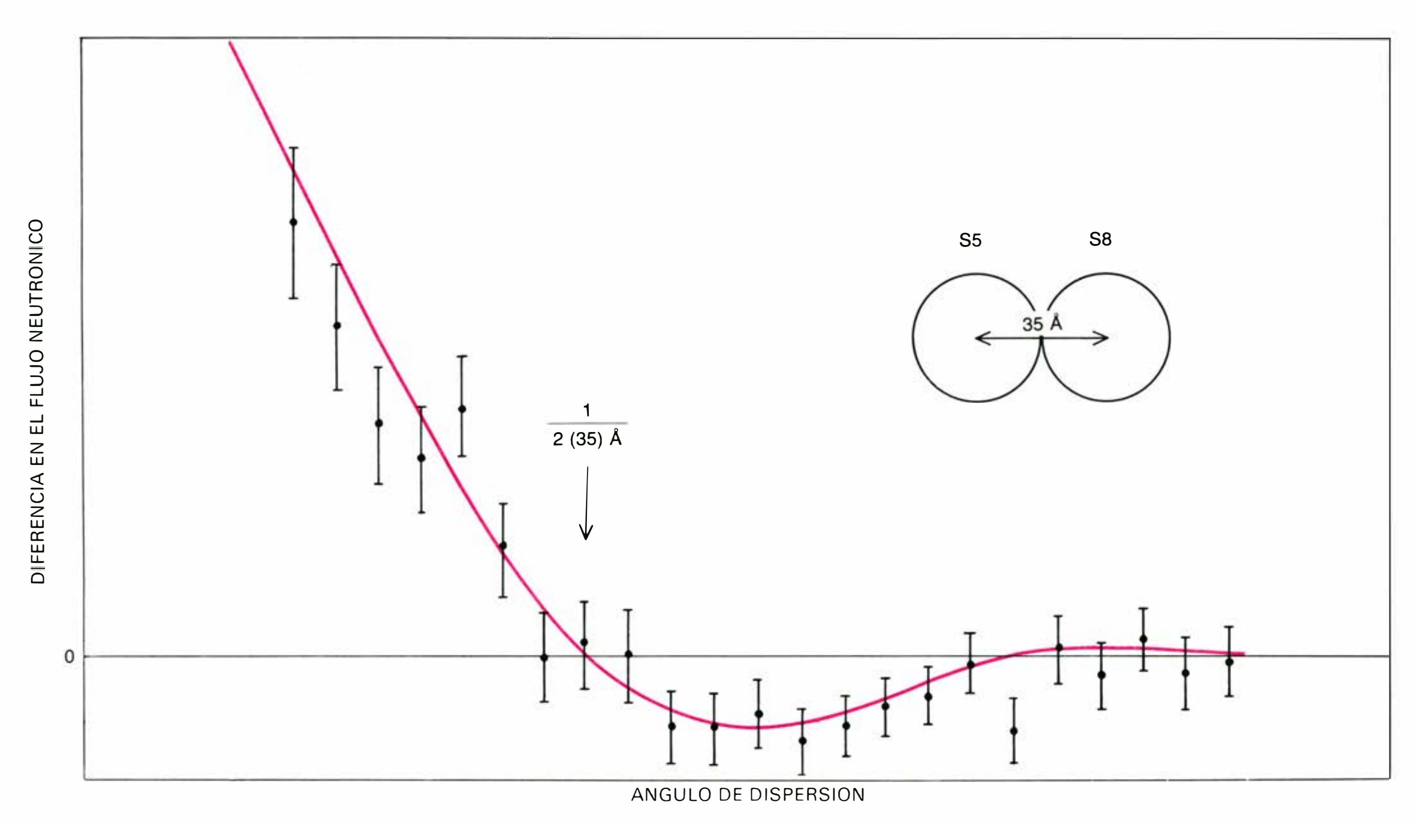
Los ribosomas que hemos empleado pertenecen a la bacteria *Escherichia coli;* nuestra primera labor fue cultivar grandes cantidades (kilogramos) de bacterias que contenían altas concentraciones de deuterio. Vimos que un procedimiento satisfactorio consistía en cultivar las bacterias en óxido de deuterio, o agua pesada, pero alimentándolas con azúcares y sales minerales normales, no deuterados. Las proteínas de estas bacterias contenían deuterio en más del 95 por ciento de los sitios normalmente ocupados por hidrógeno. El hecho de que las bacterias crecieran bien en agua pesada puso de manifiesto que la sustitución del hidrógeno por deuterio sólo causaba una ligera perturbación en el funcionamiento de las células; y todo ello parecía confirmarnos que la sustitución no alteraría la estructura de los ribosomas. Efectivamente, cuando comparamos la capacidad de intervenir en la síntesis proteica, en un sistema libre de células, por parte de los ribosomas normales y por parte de los ribosomas deuterados, observamos la no existencia de discrepancia alguna.

Una vez preparadas células normales y células deuteradas en cantidad suficiente, extrajimos los ribosomas de cada tipo celular y aislamos las subunidades menores. Luego, se desmontaron químicamente las subunidades, aislando y purificándose el ARN por separado. Mediante la técnica de cromatografía de intercambio iónico, se fueron separando una a una las 21 proteínas de la subunidad menor. En este procedimiento las proteínas quedan adsorbidas en sustrato de celulosa modificada, y se distinguen una de otra en razón de las diferentes velocidades con que son arrastradas a través del sustrato por una solución salina. En las debidas condiciones, pueden aislarse todas las 21 proteínas. Las proteínas normales y deuteradas se separan exactamente de la misma manera, lo cual vuelve a poner de relieve que cualquier alteración provocada por la deuteración no es profunda.

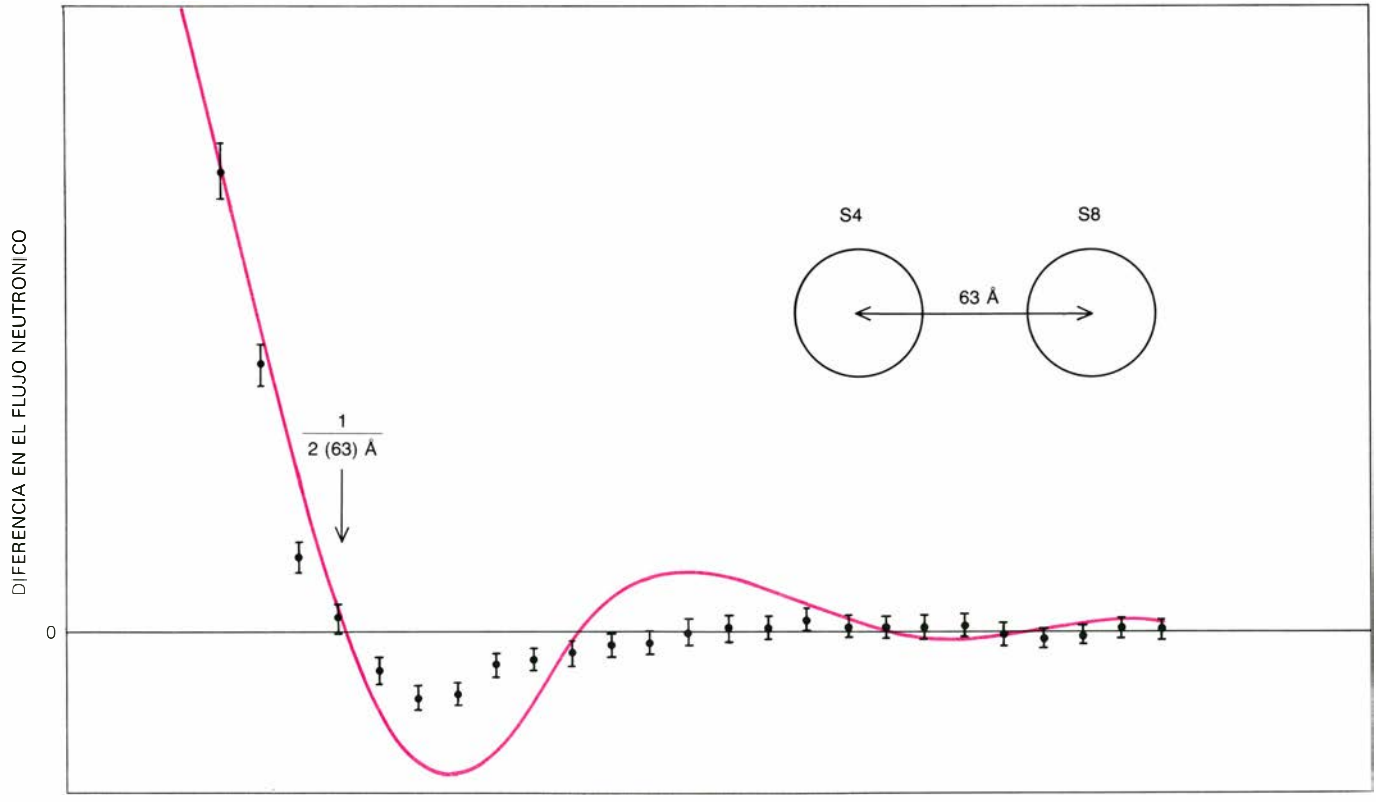
Las proteínas así purificadas se designan con las siglas *S*1, *S*2, etc. hasta *S*21 (la *S*, de *small,* denota la "unidad menor"; las proteínas correspondientes a la subunidad mayor, o *large,* se indican con las siglas *L*1, *L*2, etc. hasta *L*34). Cada una de las 21 proteínas puede identificarse por su velocidad relativa de migración a través de la columna de intercambio iónico, y también por otras características. En cualquier experimento dado se seleccionan dos proteínas; así, en uno de nuestros trabajos de laboratorio, medimos la separación entre las proteínas *S*5 y *S*8. Estas dos fracciones se seleccionaron del conjunto de las 21 proteínas hidrogenadas normales; las 19 proteínas normales restantes se mezclaron en solución. Las fracciones *S*5 y *S*8 del conjunto de las proteínas deuteradas se añadieron luego a la solución, incorporándose finalmente el ARN purificado que corresponda a las subunidades menores normales. Los disolventes empleados en la separación y purificación se substituyeron entonces por otra solución tampón y se agitó la mezcla durante dos horas a temperatura suave. En estas condiciones, las 22 moléculas componentes se replegaron y organizaron, reconstruyendo subunidades ribosómicas. El proceso está lleno de interés y aunque lo hemos presenciado muchas veces todavía sigue asombrándonos. La única información establecida explícitamente en el código genético es la que especifica la secuencia lineal de los aminoácidos en una proteína. Se acepta de manera general que esta secuencia controla también el replegamiento de la cadena de aminoácidos dando lugar a su correcta conformación tridimensional. La reconstrucción espontánea de las subunidades ribosómicas implica que el código genético contiene también la información necesaria para montar un gran complejo nucleoproteínico, de moléculas proteicas y ARN.

Las unidades reconstruidas contienen las dos proteínas deuteradas, cuya identidad química es conocida y cuyas posiciones hay que determinar. El resto de la estructura contiene únicamente hidrógeno. Las pruebas realizadas con las subunidades reconstruidas han demostrado que son activas en la síntesis proteica, que son similares a los ribosomas normales en cuanto a tamaño y forma y que las proteínas deuteradas están efectivamente incorporadas a la estructura.

Antes de medir la curva de dispersión de neutrones, es necesario hacer otra operación preparatoria que es decisiva también. Si las subunidades deuteradas estuvieran en suspensión en un medio de agua pura, el haz de neutrones



MIDIENDO LA DISTANCIA entre dos componentes de la subunidad menor, las proteínas designadas por *S*5 y *S*8, se obtiene una separación de 35 angstroms entre centros. La curva de interferencia medida viene dada por los puntos negros y las barras verticales representan el posible error experimental. La curva sólida es una teórica calculada suponiendo que las dos proteínas sean esféricas, que las dos tengan 35 angstroms de diámetro y que están tocándose. Dado que la curva teórica se corresponde bastante bien con los datos medidos, tales suposiciones parecen estar justificadas. La medición de la distancia proviene del punto donde la curva corta el cero por primera vez.



SE EMPAREJARON LAS PROTEÍNAS *S*4 Y *S*8 para medir una segunda distancia. Una vez más los puntos y las barras de error son datos medidos y la curva sólida se basa en un cálculo teórico. Pero en este caso, la curva teórica no se corresponde muy bien con los datos; en particular, la curva medida parece estar mucho más amortiguada que la calculada. Esta discrepancia indica probablemente que una de las proteínas no queda descrita correctamente cuando se la compara con una esfera y, como se vio que *S*8 era compacta al hacer las mediciones entre *S*5 y *S*8, parece probable que *S*4 tenga una forma no esférica. La desviación atribuida a la forma de *S*4 no altera sustancialmente el punto en que la curva corta por primera vez el cero, de modo que la distancia (63 A) entre los centros de las dos proteínas puede medirse todavía.

detectaría el contraste entre las proteínas con deuterio incorporado y las que tuvieran hidrógeno; pero también registraría el contraste destacado entre las subunidades tomadas como un todo y el disolvente en que están suspendidas. La señal de interferencia tiene una amplitud pequeña y es difícil de extraer incluso bajo condiciones óptimas; quedaría completamente enmascarada si no se eliminara la dispersión correspondiente al conjunto de las subunidades. El problema se soluciona con la aplicación de la técnica denominada de igualación de contrastes. Añadiendo agua pesada al disolvente, se pueden ajustar sus características de dispersión, de modo que el número de neutrones dispersados por unidad de volumen sea, en promedio, el mismo para los ribosomas que para el tampón circundante. De esta manera, el contraste entre la solución y el ribosoma considerado como un todo queda reducido prácticamente a cero y, si el ribosoma fuera un objeto completamente uniforme, desaparecería: el rayo neutrónico no detectaría su presencia. Lo cierto es, sin embargo, que el ribosoma no es uniforme; contiene dos proteínas fuertemente deuteradas, que ahora se convierten en los objetos prominentes de la solución.

Nuestras mediciones sobre la dispersión de neutrones se han realizado en el reactor de alto flujo del Laboratorio Nacional de Brookhaven. Los neutrones liberados por la fisión de núcleos de uranio en el núcleo de este reactor resultan frenados por las múltiples colisiones que sufren con los núcleos de deuterio del manto refrigerante de agua pesada que rodea el cuerpo del reactor. En nuestros experimentos, los neutrones se frenan hasta velocidades térmicas, esto es, deceleran hasta que la distribución de energías entre los neutrones es como la de las partículas de un gas en equilibrio térmico. Los neutrones con estas energías relativamente bajas son útiles en múltiples investigaciones donde se requieren resoluciones de nivel atómico, incluyendo las referentes a materiales biológicos.

Los neutrones térmicos salen del reactor por una puerta situada en el compartimiento del mismo y llegan al área experimental a través de un largo tubo que atraviesa la coraza protectora de radiaciones. Cuando los neutrones salen del reactor tienen una amplia distribución de energías, o de longitudes de onda: pero nuestras mediciones requieren un haz de neutrones tal que todos ellos tengan la misma longitud de onda. Un haz de este tipo se selecciona gracias al efecto llamado reflexión de Bragg, en honor de Sir William Bragg, quien desarrolló muchas de las técnicas que ahora se emplean en estudios de difracción de rayos X y difracción de neutrones. Los neutrones se dirigen al interior de un cristal de grafito, donde son reflejados por los numerosos planos atómicos internos. El ángulo de reflexión depende de la longitud de onda de los neutrones y, en consecuencia, el haz de neutrones se abre dando lugar a un espectro; de éste se selecciona un haz monocromático integrado por neutrones con una determinada longitud de onda. El espectro de neutrones se corresponde directamente con el espectro óptico; e igualmente, los planos de átomos responsables de la reflexión de Bragg son análogos a las rayas de una retícula para difracción óptica.

En nuestros experimentos, la orientación del cristal de grafito se ajusta de forma que queden seleccionados los neutrones con una longitud de onda de 2,37 angstroms. El haz seleccionado se dirige a lo largo de un tubo de unos dos metros de longitud, en cuyo recorrido hay varios diafragmas de pequeña abertura. De este modo el rayo es colimado y dirigido hacia el interior de la célula de cuarzo que contiene la muestra de subunidades ribosómicas.

No nos hubiera acompañado el éxito en estos experimentos si no hubiera sido por la experta y entusiasta ayuda del equipo de Brookhaven. En particular, Benno P. Schoenborn, el primer investigador que empleó la difracción de neutrones para la elucidación de estructuras biológicas, ha contribuido grandemente en el desarrollo de las técnicas necesarias. Entre sus contribuciones esenciales se cuenta la puesta a punto de un detector de neutrones capaz de medir simultáneamente el flujo neutrónico en muchos puntos de un plano, es decir, un detector de posición con buena resolución espacial.

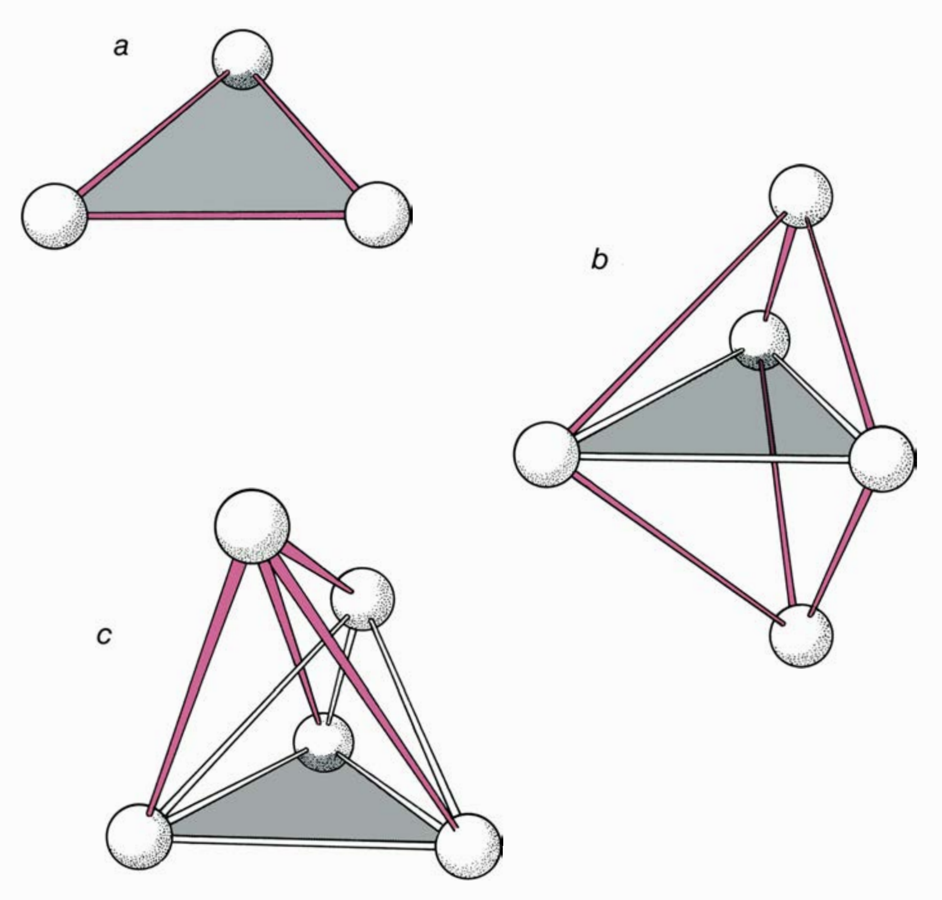
Como los neutrones no interaccionan con los electrones, no pueden ser detectados de la misma manera que una radiación electromagnética o un haz de electrones. Una película fotográfica, por ejemplo, no resulta impresionada eficazmente por los neutrones. El detector que hemos empleado en Brookhaven debe su sensibilidad a la especial propiedad que tiene el gas helio 3, el isótopo del helio con dos protones y un neutrón. Cuando un neutrón choca con un núcleo de helio 3, los productos de la colisión son un núcleo de tritio (el isótopo del hidrógeno con un protón y dos neutrones) y un protón; en efecto, el núcleo de helio 3 absorbe un neutrón y emite un protón. El protón, claro está, es portador de una carga eléctrica, la cual ioniza momentáneamente algunos de los átomos de helio vecinos. Esta ionización puede detectarse por métodos convencionales.

El detector es una cámara plana llena de helio 3. Dentro de la cámara hay una serie de electrodos dispuestos de manera que puedan detectar la ionización de cualquier pequeña región del gas en forma de impulso eléctrico. A partir de la naturaleza de la señal producida se puede determinar la posición del fenómeno de ionización. Se cuentan los impulsos, se codifican con el fin de señalar la posición del neutrón incidente y se registran en cinta magnética.

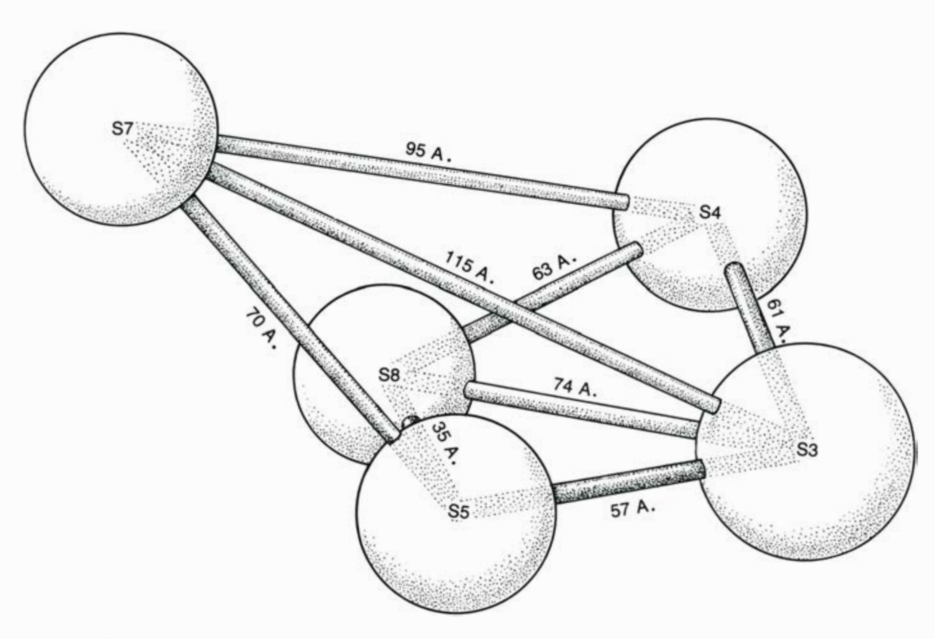
La amplitud de la señal de interferencia generada por las proteínas deuteradas es sólo un pequeño tanto por ciento del total del flujo de neutrones que inciden sobre el detector. Como resultado de ello, la estabilidad y fiabilidad del aparato experimental son críticas. Para las muestras con mayor número de ribosomas que podemos preparar fácilmente, se requieren unos tres días de mediciones continuas al objeto de recoger los datos suficientes y definir una sola curva de interferencia.

Una vez recogidos los datos, se procesan luego con la ayuda de un computador. En primer lugar, las mediciones tomadas en todas las zonas del detector se suman por anillos que tienen como centro la posición del rayo no desviado. Cada anillo, representa, por tanto, una medición del flujo de neutrones que corresponde a un ángulo de dispersión concreto. El conjunto de anillos da lugar a una gráfica del flujo de neutrones con respecto al ángulo de dispersión.

El objeto de las mediciones, claro está, no es el flujo total de neutrones, sino la curva de interferencia ondulante. Esta curva puede ser aislada sustrayendo, del flujo total medido, todos los componentes que no surgen de la dispersión provocada por la pareja de proteínas. Se hacen mediciones en dos soluciones. En una de ellas, la mitad de los ribosomas contienen la pareja seleccionada de proteínas deuteradas (tales como la *S*5 y la *S*8); el resto de ribosomas de esta solución contienen únicamente proteínas hidrogenadas. En la otra solución, la mitad de los ribosomas tienen una de las dos proteínas deuteradas (por ejemplo la *S*5) y la otra mitad de los ribosomas contienen la otra proteína deuterada (*S*8). Si la concentración de las dos soluciones es igual, ambas contienen entonces los mismos compo-



LA ESTRUCTURA TRIDIMENSIONAL de la subunidad menor se obtiene por triangulación a partir de las mediciones de las distancias entre pares de proteínas. Las tres primeras mediciones definen un triángulo y, necesariamente, un plano a través de la subunidad *(a)*. Otras tres mediciones *(b)* dan la posición de una cuarta proteína y definen un tetraedro. Al situar esta cuarta proteína se presenta, sin embargo, una ambigüedad que no puede ser resuelta por ningún sistema basado en pares de mediciones: las proteínas podrían ocupar cualquier posición a cada lado del plano definido por el triángulo original. Si se escoge arbitrariamente, se pueden añadir otras proteínas a la estructura, midiendo cuatro proteínas cualesquiera cuyas posiciones estén ya definidas *(c)*. No obstante, la simetría de la estructura permanece indeterminada, es decir, el mapa podría representar la estructura real o su imagen especular. Un mapa de la subunidad requiere unas 74 mediciones.



ESTE MAPA BOCETO de una parte de la subunidad ribosómida menor comprende cinco proteínas relacionadas por ocho mediciones de distancia. En el mapa no se representa un tetraedro completo, y por ello la estructura figurada no es rígida, sino que puede doblarse según diversos ejes, cambiando la orientación de las proteínas sin cambiar las distancias establecidas entre ellas. La medición entre *S*5 y *S*7 es solamente a título de ensayo. Las proteínas están representadas por esferas con el fin de indicar sus volúmenes aproximados, aunque algunas no sean esféricas.

nentes a saber: el mismo número de ribosomas y el mismo número de proteínas deuteradas de cada tipo. Difieren solamente en la disposición de estos componentes. Si la curva de la intensidad de dispersión de una solución se sustrae de la de la otra, la diferencia es la parte correspondiente a la distribución geométrica de las proteínas en los ribosomas doblemente deuterados. Esta es la curva de interferencia, la equivalente a las bandas observadas en el experimento óptico con dos hendeduras.

En 1974, cuando obtuvimos nuestras primeras mediciones de la distancia que separaba un par de proteínas, quedamos desolados al encontrar que la curva de interferencia ondulante que habíamos esperado tenía una única fluctuación; en otras palabras, la intensidad del flujo neutrónico iba de un alto valor positivo a un valor negativo, volviendo luego a cero, sin posteriores oscilaciones. Al revisar nuestro modelo teórico de la curva, volvimos a tranquilizarnos: el modelo simple presuponía que las proteínas eran esféricas o por lo menos compactas, y éste no es necesariamente el caso. Si las proteínas tuvieran una forma menos regular, podía dar como resultado una fuerte amortiguación de los picos subsiguientes en la curva. La distancia desde el origen al punto donde la curva corta el cero por primera vez, sin embargo, queda poco afectada por la forma de la proteína, de modo que la distancia entre las moléculas podía medirse aún. Con el fin de asegurar que la señal que habíamos detectado era real, repetimos el experimento, siguiendo paso a paso todo el proceso, pero empleando únicamente proteínas hidrogenadas; no se detectó ninguna señal de interferencia. Otros procedimientos de control confirmaron que la señal observada está generada por las proteínas deuteradas y que las subunidades ribosómicas sobreviven a las condiciones experimentales.

Además, pudimos repetir las mediciones obteniendo resultados razonablemente coherentes.

En una primera aproximación, la distancia entre las proteínas puede obtenerse por un procedimiento simple. Hemos visto que la separación entre las franjas de interferencia es inversamense proporcional a la distancia entre las proteínas; esta relación puede convertirse fácilmente en una expresión cuantitativa. En primer lugar, se mide el ángulo entre el haz no desviado y el borde de la franja de interferencia más interna (donde la curva de interferencia corta el cero por primera vez). La distancia entre

las proteínas es entonces igual a la longitud de onda del neutrón dividida por el doble del seno de dicho ángulo.

Aplicando esta técnica a la curva de interferencia correspondiente a las proteínas *S*5 y *S*8, da como resultado una distancia entre centros de unos 35 angstroms. Para este par resulta totalmente adecuado un modelo muy simple sobre la forma de las proteínas. Considerando que ambas proteínas son esferas con volúmenes que concuerdan con su peso molecular y considerando que están en contacto, se puede calcular una curva de interferencia que se ajuste bien a los datos. Por ello concluimos que *S*5 y *S*8 son compactas cuando se observan según la visual que pasa por sus centros de gravedad y que están a unos 35 angstroms una de otra. Dado su peso molecular, esta es la distancia mínima a que pueden estar situadas.

Otro par de proteínas que hemos medido es la *S*4 y la *S*8. Aquí el lugar donde se intercepta el cero por primera vez da una distancia interproteica de 63 angstroms; pero una curva calculada para dos proteínas esféricas de tamaño apropiado no se ajusta bien a los datos. Por ello parece que los efectos derivados de la forma de la proteína han alterado la forma de la curva. El modelo matemático sugiere que se generaría una curva de interferencia del tipo correcto si una de las proteínas fuera compacta y alargada la otra. Como *S*8 parece ser compacta en la medición *S*5-*S*8, es posible que *S*4 sea alargada.

Una vez obtenidas unas cuantas distancias entre pares de proteínas, cabe empezar la construcción de un mapa de la subunidad ribosómica menor. El mapa empieza con tres mediciones centro a centro, relacionando tres moléculas proteicas; estas tres definen necesariamente un triángulo y un plano en la subunidad. Puede localizarse una cuarta proteína calculando la distancia entre ella y las tres primeras; determinándose así el vértice de un tetraedro, a no ser que esté situada en el mismo plano.

La situación en el mapa de la cuarta proteína comporta una cierta ambigüedad, difícil de salvar. Las distancias medidas entre las tres primeras definen de hecho dos posibles posiciones para la cuarta, una a cada lado del plano; los datos obtenidos por el método de dispersión de neutrones no contienen ninguna información acerca de cuál sea la posición correcta. Si se hace una elección arbitraria, entonces las restantes proteínas que se añadan al mapa pueden situarse sin ambigüedad, pero persiste la cuestión de que podría representar o bien la estructura real o bien su imagen especular. Esta ambigüedad respecto a la simetría de la estructura no puede ser resuelta por ningún sistema de medición entre pares de proteínas, y así la ambigüedad permanecerá, incluso cuando el mapa esté completado. Sin embargo, si llegáramos al punto en que la simetría del ribosoma fuera la única incógnita en nuestro conocimiento sobre su estructura, habríamos conseguido mucho.

Una vez establecidas las posiciones correspondientes a cuatro proteínas, que definen un tetraedro, pueden añadirse al mapa nuevas proteínas, haciendo un conjunto de cuatro mediciones respecto a los vértices de este tetraedro o respecto a otras cuatro moléculas cualesquiera cuyas posiciones estén completamente definidas. No es necesario medir todas las distancias interproteicas para desarrollar un mapa completo de la estructura; para las 21 proteínas de la subunidad menor hay 210 pares proteicos, pero son suficientes 74 mediciones para definir la estructura. Evidentemente, las mediciones adicionales aumentarían la seguridad y la fiabilidad del mapa.

La mayoría de las mediciones que hemos obtenido hasta aquí se hicieron en colaboración con Jerome Langer, estudiante postgraduado en la Universidad de Yale. El mapa de la subunidad menor que hemos reconstruido consta de cinco proteínas: las *S*3, *S*4, *S*5, *S*7 y *S*8. Entre ellas se han determinado cuatro triángulos mediante ocho mediciones, pero aún no se conocen definitivamente las relaciones espaciales de los triángulos entre sí. Una medición adicional a modo de ensayo sugiere que los triángulos forman un agregado plegado y relativamente compacto; pero incluso si se demuestra que esta medición es correcta, no determinará de manera rígida las relaciones espaciales de las cinco proteínas. Se necesita, por lo menos, otra medición para definir un tetraedro.

Claro está que con sólo ocho distancias interproteicas establecidas y un mínimo de 74 mediciones, el análisis estructural de la subunidad menor del ribosoma acaba de empezar. Hemos tardado un año en adquirir la información que ahora poseemos, y serán necesarios unos cuantos más para completar el mapa. Nuestro avance puede parecer lento, pero es similar al registrado en otros muchos tipos de empresas científicas; además, a causa de la importancia de los ribosomas en el metabolismo celular, son grandes los dividendos en potencia si se tienen en cuenta los nuevos conocimientos.

Un mapa que revele las posiciones relativas de las 21 proteínas no es el único producto esperado del estudio de las curvas de interferencia con neutrones. Como hemos apuntado, la forma de la curva puede dar más información acerca de la morfología y la orientación general de cada proteína. En principio, el ARN ribosómico podría incluirse también en el mapa de difracción por neutrones. Para aplicar esta técnica al ARN, sería necesario un método para cortar su cadena en puntos conocidos a lo largo de la misma, introduciendo luego segmentos deuterados en la molécula reconstruida.

Durante la síntesis proteica interaccionan con el ribosoma varias proteínas y otras moléculas, entre las cuales está el ARN mensajero y el ARN de transferencia.

Y si cualquiera de estas interacciones fuera estable, o si pudiera detenerse en alguna forma estable intermedia, el locus de interacción podría determinarse relacionando estas moléculas con el mapa proteico. Asimismo, podrían medirse las distancias interproteicas dentro del ribosoma durante algunas fases de la síntesis proteica, considerando, una vez más, que se formaran estados intermedios estables. Durante la síntesis proteica, la configuración del ribosoma puede cambiar; las mediciones por difracción de neutrones revelarían la naturaleza de estos cambios. La información obtenida a partir de estudios de este tipo nos introduciría en la comprensión de uno de los procesos céntricos de la vida: el mecanismo por el cual el código genético se traduce en expresión fenotípica.

Por último, el ribosoma no es la única estructura biológica que puede explorarse provechosamente mediante dispersión de neutrones. Un agregado particularmente interesante, que se espera sea pronto estudiado, es la ARN polimerasa dependiente de ADN, que interviene en el proceso de la transcripción copiando el código del ADN en forma de ARN mensajero. Otros sujetos de investigación por el mismo método son la cromatina, el complejo molecular que incluye el ADN, el cual constituye el material genético de los organismos eucariotas, y las distintas membranas celulares. Una excitante perspectiva de comprensión, a nivel molecular, de todas esas estructuras.

Ciencia y sociedad

Energía nuclear (3)

Los procesos de transmutación atómica son de dos tipos. En los de fisión, se produce la fragmentación de un núcleo pesado, como el del uranio, mediante su bombardeo con neutrones, originándose dos núcleos de elementos más ligeros, con liberación de enormes cantidades de energía, del orden de 2×10^8 electrón-voltios. Los procesos de fusión, por el contrario, consisten en la combinación de dos o más núcleos atómicos ligeros para formar otro más pesado, reacción que también es exotérmica, es decir, liberadora de energía.

A partir de un cierto umbral, las reacciones de fisión se sustentan a sí mismas, ya que en ellas se liberan nuevos neutrones, que provocan la fisión de nuevos núcleos, hasta que el proceso afecta a todo el material fisionable que se encuentra presente. Es lo que se conoce como reacción en cadena. En el caso de la fusión, la interacción de los núcleos es función directa de la temperatura y, por tanto, una vez iniciada, el aumento de temperatura que origina favorece también su continuación. En suma, ambas reacciones son, en determinadas condiciones, explosivas. Las bombas arrojadas sobre Hiroshima y Nagasaki al finalizar la segunda guerra mundial fueron de fisión nuclear, mientras que la bomba termonuclear o bomba H, conseguida en 1952-1953 se basaba en la fusión.

La utilización de la energía nuclear como fuente energética para usos pacíficos requiere no sólo que los procesos puedan mantenerse una vez iniciados, sino también que se les pueda gobernar, es decir, que se evite su desarrollo incontrolado que terminaría en explosión. Las instalaciones para el desarrollo controlado de ambos procesos son los denominados reactores nucleares.

Los reactores industriales actualmente existentes en gran número de países utilizan todos el mecanismo de la fisión. La fisión del uranio es controlada mediante la utilización de diversos productos, llamados moderadores, que absorben neutrones y mantienen el ritmo de la reacción dentro de los límites establecidos. El calor generado se transmite por medio de líquidos o gases de las turbinas que producen electricidad. El combustible utilizado en los reactores es el uranio, pero el uranio natural está compuesto de dos isótopos, el uranio-238 y el uranio-235. Sólo este último permite la reacción en cadena y su presencia en el uranio natural es de aproximadamente un átomo por cada 140 de uranio-238. Por tanto, a efectos de su utilización como fuente de energía a gran escala, el uranio presenta dos limitaciones fundamentales: como recurso natural que es, las cantidades disponibles sobre el planeta Tierra son limitadas; en segundo lugar, el bajo porcentaje de rendimiento del uranio empleado actualmente origina un despilfarro innecesario de este recurso escaso. A largo plazo, las perspectivas del proceso de fisión nuclear están condicionadas por la puesta a punto de los llamados reactores regeneradores o nodriza, cuyo funcionamiento produce más material fisionable que el que consumen. Aunque desde 1970 se está realizando un gran esfuerzo investigador en diversos países sobre este tipo de reactor, hasta la fecha no existen más que prototipos y el tiempo requerido para duplicar la cantidad de combustible consumido por el reactor es todavía de más de 10 años. La vía de la fusión no tiene estas limitaciones. Los elementos fusionables se encuentran disponibles en cantidades prácticamente inagotables. Sin embargo, la técnica correspondiente está sólo en sus primeros balbuceos. En palabras de H.A. Bethe, premio Nobel de Física e indiscutible experto en la materia: "Todavía ignoramos por completo si el proceso de fusión podrá producir nunca energía utilizable. Es cierto que las estrellas y las bombas de hidrógeno obtienen su energía de la fusión de núcleos atómicos ligeros, pero ¿podrá desencadenarse esta energía sobre la tierra de forma controlada? Los requisitos para lograrlo son enormes: una mezcla de gases de hidrógeno pesados debe ponerse a una temperatura de unos 100 millones de grados, temperatura que ha de mantenerse durante el tiempo necesario para que se produzcan las reacciones entre los núcleos de hidrógeno a un ritmo suficiente para originar un superávit neto de energía utilizable". No se cree que antes de 1980 puedan tenerse datos decisivos que permitan elegir entre las diversas propuestas técnicas existentes para resolver estos problemas. Pero, incluso en este caso, de la fase de diseño a la de realización industrial media un largo trecho, por lo que no es verosímil que hasta el año 2000 puedan funcionar prototipos de reactores comerciales de fusión, y para 2020 la fusión no podría contribuir todavía más que con un porcentaje muy pequeño a las necesidades energéticas mundiales.

Vemos, pues, que ni siquiera la energía nuclear constituye la panacea que popularmente se piensa para la resolución del problema energético, lo que ayuda a comprender la gravedad de este último. Dentro del campo nuclear, la energía obtenida por fisión es la única vía practicable actualmente, y eso con grandes limitaciones. Además, tiene peligros, de los que nos ocuparemos el mes próximo.

Potenciales de producción

Entre las necesidades más conspicuas del hombre actual, y ya no sólo como proyección al futuro, se encuentra la de incrementar la producción primaria a sus niveles máximos potenciales. Evidentemente, son múltiples y diversas las vías mediante las cuales pueden obtenerse mayores rendimientos en la obtención de la biomasa vegetal capaz de canalizar el flujo energético desde el sol hasta los niveles tróficos superiores, en los que se halla integrada la especie humana.

Refiriéndonos concretamente a las plantas cultivadas, componentes directos o indirectos de nuestra dieta, los métodos generales a utilizar se concretan en: *a*) la reducción de la diversidad agrícola en favor de las especies más productoras; *b*) la selección genética de las variedades cultivadas en orden a una mayor productividad, adaptación climática, resistencia a enfermedades, enriquecimiento proteico o vitamínico, etc.; *c*) la incorporación, mediante episomas, plásmidos o mutaciones dirigidas, de características principales como fijación del nitrógeno elemental o metabolismos de alto rendimiento y *d*) la modificación de parámetros ambientales, nutricionales y otros diversos para desarrollar al máximo las potencialidades endógenas de una determinada especie o variedad.

Bajo la dirección de uno de los autores, se ha concretado precisamente en este último aspecto la actuación del Departamento de Fisiología Vegetal de la Universidad de Barcelona, realizada en los

Campos Experimentales de Pedralbes (Barcelona) sobre praderas mediterráneas artificiales de siega continua. Los esfuerzos se han dirigido hacia dos contextos: la obtención de producciones totales que alcancen los máximos teóricos calculados y el logro de una mayor uniformidad de distribución de la biomasa recolectada respecto a las variaciones estacionales.

La distribución del crecimiento de las especies pratenses a lo largo del año ofrece un marcado carácter estacional. Ello es debido a las características genéticas de cada especie o variedad que condicionan comportamientos diferentes con respecto a los factores climáticos variables con la sucesión de las estaciones. A este respecto, con las principales especies cultivadas pueden establecerse dos grandes grupos. Uno, que contiene la mayor parte de las mismas, se caracteriza por presentar óptimos térmicos próximos a 20° C, luminosidades para la asimilación fotosintética óptima menores que la solar y requerimientos hídricos elevados *(Lolium, Dactylis, Festuca)*; otro, mucho más reducido, que precisa óptimos términos entre 30 y 35° C, luminosidades máximas y menores exigencias de agua *(Cynodon)*.

El clima mediterráneo, con amplias variaciones estacionales, inviernos fríos, primaveras y otoños suaves y pluviosidad suficiente y veranos calurosos, luminosos y secos, obliga a que, en general, las especies del primer grupo encuentren solamente un ambiente apropiado en primavera y principios de verano y otoño, y las del segundo, en la época estival. Así, la distribución del crecimiento en una planta del primer grupo tiene carácter bimodal, con dos óptimos de crecimiento separados por la depresión estival y la inhibición invernal. En el caso de una especie del segundo grupo el perfil ofrece un máximo principal centrado en el verano, con un período vegetativo corto. Quedan, pues, complementadas dentro del año las producciones de ambos tipos de pratenses. Aunque el sistema de cultivo mixto presenta características muy interesantes, y con determinados fines se utiliza en pratidultura, no resulta el más adecuado si se desean estudiar las capacidades máximas de producción de una especie determinada.

Las pratenses más interesantes para climas templados se encuentran entre las del primer grupo, ya que las otras, con máximos de producción estival de mayor rendimiento por superficie cultivada, tienen una duración tan escasa del período vegetativo que su producción total se ve mermada. Entre los fines que se persiguen en un cultivo de pratenses de clima templado destacan la reducción de las variaciones estacionales en la producción, el aumento de la producción total y la consecución de un producto vegetal de mejor calidad (digestibilidad, sapidez, etc.). En cuanto a las producciones anuales y reales de forraje seco en praderas cultivadas de zona templada, se considera ya a las cifras de 15 tm por hectárea y año como rendimientos muy elevados.

La proporción entre la energía luminosa aprovechable en la fotosíntesis y la que es almacenada en la materia orgánica seca queda entre el 1 y el 2 por ciento. Durante cortos períodos, dentro de los máximos de producción, puede llegarse al 4 o 5 por ciento pero, en cambio, durante las depresiones, se llega en ocasiones a valores nulos.

Estas producciones elevadas se consiguen con elaborados programas de cultivo, irrigaciones frecuentes de los pastizales, regímenes de cosechado seleccionados y dosis fuertes de abono, sobre todo nitrogenado. Así se adelanta el inicio del crecimiento, se reducen las depresiones y se suele aumentar la calidad.

Varios autores han calculado las producciones máximas potenciales y las cifras que se obtienen son del orden de las 30 a 35 tm, ha^{-1} $año^{-1}$ es decir, el doble de las buenas cosechas reales.

Con el fin de aproximar los rendimientos actuales a los máximos potenciales se proponen varios caminos: aumentar las eficiencias de conversión de energía luminosa (luz-materia seca), alargar los períodos de producción, sobre todo llenando las depresiones, etc. Todo ello, por supuesto, sin recurrir a metodologías excepcionales u onerosas y simplemente conociendo y regulando los procesos fisiológicos que intervienen en la síntesis de la materia orgánica.

En nuestros ensayos se ha encontrado como factor primordial, la irrigación frecuente, sobre todo y como es lógico durante la estación seca y calurosa. Así, la transpiración se mantiene activa y la temperatura de las hojas y de las raíces queda más baja y no tan alejada de sus óptimos térmicos.

Otro de los parámetros que intervienen significativamente en la producción de las praderas de cosecha continua es la cadencia de los cortes. Dadas las características sigmoidales de las curvas generales de crecimiento es necesario hallar, para cada variedad ensayada y zona climática determinada, una intermitencia tal que sea suficientemente amplia como para alcanzar la fase de crecimiento rectilíneo máximo y hacer mínimo el gasto de energía mecánica de recolección por equivalente de energía alimenticia obtenida, y suficientemente frecuente como para que el vegetal no llegue a entallar, lo que provocaría una pérdida apreciable en la producción total. En nuestras praderas, el período de corte ha quedado establecido en 21 días, con lo cual, además, se llega a obtener, en las épocas de máxima productividad, un LAI (índice foliar o índice de cobertura de la planta respecto al suelo que ocupa) entre 3 y 5 metros cuadrados hoja m^{-2} suelo, que se reconoce óptimo para un mejor aprovechamiento de la luz solar.

El factor nutricional más limitante en este tipo de praderas lo constituye, sin duda alguna, el aporte de nitrógeno mineral, del cual hemos estudiado no solamente los efectos dosis/producción (que son lineares hasta concentraciones altas, del orden de 1500 Kg N ha^{-1} $año^{-1}$) sino también la repercusión estacional respecto a la forma de suministro nítrico o amoniacal. Hay que advertir, sin embargo, que estas dosis tan altas se justifican solamente por los riesgos, intensos que irremediablemente provocan importantes pérdidas por lavado. En este sentido hemos podido averiguar como, sin originar prácticamente repercusiones en la linearidad de respuesta, las altas concentraciones (1000 kg N ha^{-1} $año^{-1}$) de abonado nitrogenado en forma catiónica (NH_4^+) y las bajas concentraciones (250 kg N ha^{-1} $año^{-1}$) de abono aniónico (NO_3^-) aumentan significativamente el rendimiento, mientras que dosis intermedias (500 kg N ha^{-1} $año^{-1}$) determinan producciones que difieren significativamente con ambos nutrientes. Además, en los meses de mayor irradiación solar se manifiesta una tendencia benéfica con el abonado en forma de nitrato (¿en relación con la fotorrespiración?) respecto a la fertilización amoniacal, situación que se hace opuesta en la época de irradiación baja. Si a todo ello sumamos el hecho de que en la depresión estival se produce una inversión de la relación dosis/efecto, queda bastante claro que existen diversos puntos de actuación para la regulación estacional de las producciones, mediante el ensayo de diversas proporciones y dosis de los iones amonio y nitrato.

Con estas normas especializadas de cultivo, aplicadas en praderas de *Lolium perenne* y mantenidas durante muchos años sin necesidad de resiembra alguna, hemos logrado cosechas de siega del orden de 20 tm de peso seco ha^{-1} $año^{-1}$, sobrepasándose, en ocasiones las 23 tm ha^{-1} $año^{-1}$. Las eficiencias medias de conversión de energía luminosa en materia orgánica ha sido próximas al 2 por ciento y con máximos durante las cimas

de crecimiento del 4 y 5 por ciento, excepcionalmente, en algún primer año del cultivo, por encima del 6 por ciento.

En estos cultivos de raigrás desaparecen las depresiones acentuadas del crecimiento y los perfiles anuales de la distribución de la tasa de producción quedan aplanados y de curso ondulado o multimodal. Si se incluye la producción neta que supone el rastrojo remanente de las siegas, se deben sumar como mínimo de 3 a 4 tm más cada año. También es necesario tener presente que durante el período vegetativo anual la producción de raíz supone otra cifra importante que eleva en realidad la neta total en 4 a 5 tm. Estas cifras son bajas debido al riesgo, al abandono y a los cortes aplicados con elevada frecuencia. De todo ello resulta una producción real mínima del orden de 30 a 32 tm de peso seco ha^{-1} $año^{-1}$.

Aunque las posibilidades para mejorar los rendimientos no han quedado agotadas, las producciones netas conseguidas son muy próximas ya a las potenciales máximas calculadas (A.C. y F.G.)

ADN recombinante

Diversos organismos científicos elaboran las normas generales que regulan el uso de las técnicas que se utilizan en "ingeniería genética".

Pocas técnicas científicas han provocado tantas discusiones y han aglutinado tanta especulación como las que hoy se utilizan en algunos laboratorios de biología molecular para unir artificialmente fragmentos de ADN de origen distinto formando así una molécula de ADN recombinante. Mediante estas nuevas técnicas que permiten el desarrollo de lo que hoy se conoce como "ingeniería genética" (esto es, la introducción de una planificación humana en la formulación de nuevos genes y nuevas combinaciones de genes), y con el concurso de los enzimas de restricción, los científicos pueden obtener genes aislados de organismos complejos e insertarlos después en el relativamente simple genoma de una bacteria, generalmente *Escherichia coli,* donde se podrá estudiar la expresión del gen aislado mediante el análisis de las proteínas que sintetiza. Aunque hoy tenemos datos suficientes para pensar que existen incompatibilidades serias entre los mecanismos de expresión génica de los organismos sencillos y complejos, que deberán ser superadas con esfuerzo y trabajo considerable, los científicos y la sociedad en general han mostrado su preocupación por las secuencias que podrían derivarse de unos experimentos que, de hecho, permiten traspasar las barreras entre distintas especies las cuales se han establecido fruto de un proceso evolutivo a lo largo de millones y millones de años.

Las técnicas de obtención de ADN recombinante implican el uso de unos enzimas que permiten fragmentar el material hereditario de organismos vivos, reconociendo secuencias específicas de ADN, con una precisión hasta ahora insospechada. Puede separarse así un fragmento de ADN portador de uno o varios genes del resto de cromosoma para ser posteriormente insertado en otro segmento de ADN de un organismo distinto. Estos enzimas reciben el nombre de "enzimas de restricción" y existen en bacterias precisamente para defenderlas de un posible intercambio de material genético entre distintas especies. Su función es, pues, en la célula bacteriana, la de reconocer cualquier ADN extraño y fragmentarlo para impedir su expresión. Dada su abundancia (se encuentra en aproximadamente (1/3 de las cepas bacterianas analizadas) y la precisión con que actúan (cada enzima reconoce una secuencia determinada de ADN), los biólogos moleculares poseen hoy un material valiosísimo para la determinación de las secuencias de nucleótidos de un ADN.

Los ingenieros genéticos que hacen uso de las técnicas de obtención de ADN recombinante, que ellos mismos describen como "revolucionarias" y "uno de los avances más importantes de la Biología del siglo XX", y otros científicos que no trabajan en este campo afirman que, si bien la inserción de genes extraños en un organismo podría ser la causa de que éste se comportara de manera extraña, desde luego impredecible y por tanto peligrosa, estas técnicas pueden aportar beneficios muy positivos para la sociedad en general. Se ofrece por vez primera al hombre un nuevo poder frente a la naturaleza: el poder de intervenir en la evolución, programando y creando nuevas combinaciones de genes de distinto modo y desde luego mucho más rápido de lo que ocurre en la naturaleza.

El campo de aplicación de estas técnicas sería doble: la producción y el análisis. En el primero se trataría de introducir en una bacteria genes que codifican para proteínas comercialmente importantes, como, por ejemplo, la insulina y obtener así una fábrica viviente de proteínas que son difíciles de obtener de otro modo y por tanto muy costosas. Las aplicaciones prácticas podrían ser múltiples y entre ellas se mencionan además el dotar a las plantas de cultivo con genes fijadores de nitrógeno para ahorrar el empleo de fertilizantes, construir microorganismos capaces de sintetizar productos que actualmente obtenemos del petróleo, etc. Y en el campo analítico, aunque su proyección no esté en principio tan clara, se favorecería el estudio de la organización estructural y molecular de genes de organismos superiores, incluyendo el hombre. Aunque las dos finalidades difieren considerablemente los medios necesarios para su consecución no están lejanos.

Si bien la mayoría de científicos tienden a considerar que el desarrollo de estas técnicas, al aportar grandes avances respecto la manipulación del material hereditario, nos permitirían conocerlo mejor y por tanto alcanzar un cierto dominio sobre él repercutiendo finalmente en beneficio de la sociedad, lo cierto es que hoy por hoy las técnicas ofrecen en potencia una mezcla espectacular de grandes beneficios y enormes peligros. Entre los experimentos que se consideran más peligrosos se encuentra el denominado "shotgun" (perdigonada) que consiste en fragmentar el genoma completo de un organismo en segmentos de aproximadamente unos pocos genes cada uno, unir cada uno de estos fragmentos covalentemente a una molécula portadora o vector (generalmente un fago o un plásmido) que introduce esta molécula recombinante en un huésped bacteriano adecuado (el más usado es *E. coli*), y establecer un cultivo para clonar la bacteria y obtener así cantidades suficientes de los nuevos productos génicos para el análisis. Muchos investigadores piensan que estos experimentos de "perdigonada" encierran un gran peligro en potencia dado que el genoma total de un organismo contiene genes que podrían ser letales al hombre si se expresan en *E. coli,* que vive frecuentemente en el intestino humano.

Preocupados por los posibles daños que pudieran derivarse del uso de estas técnicas un comité formado por científicos de la Academia Nacional de Ciencias de los Estados Unidos, presidido por Paul Berg, de la Universidad de Stanford, pidió en julio de 1974 a todos los científicos comprometidos en este campo que parasen momentáneamente sus experimentos en espera de que una conferencia internacional decidiera sobre los riesgos implicados y reglamentara cómo y en qué condiciones debían llevarse a cabo. La conferencia propuesta tuvo lugar en febrero de 1975 en Asilomar (Estados Unidos) con participación de científicos de todo el mundo. Allí se establecieron las normas básicas para

seguir los experimentos de recombinación que debían servir de referencia o de modelo a las que posteriormente adoptase cada país. El tiempo de espera, que se había prolongado en exceso, y los distintos puntos de vista de cada investigador, que diferían si trabajaba o no con ADN recombinante, hicieron que algunas de aquellas charlas fueran realmente explosivas. Se rumoreaba, además, que algunos científicos, cuya investigación se había limitado exclusivamente a este campo, realizaban los experimentos los sábados por la tarde con el riesgo que esto suponía no sólo para ellos mismos sino para la población en general. Afortunadamente, pues, se acordó redactar unas normas básicas de trabajo a las que deberían ajustarse todos los científicos mientras no se elaborasen unas propias para cada país.

Entre las innumerables sugerencias de Asilomar hubo quien propuso la prohibición de realizar experimentos en los que se utilizara el genoma de *E. coli* para expresar los nuevos genes añadidos. Esto implicaba sacrificar el organismo mejor conocido a nivel genético en los experimentos de recombinación, en virtud de que éste no es extraño al cuerpo humano. La postura opuesta se basaba en que precisamente *E. coli* era el organismo más adecuado frente a otros muchos que, además de no poder predecir su utilidad y que por ser menos conocidos resultaban todavía más peligrosos, estaban sólo amparados por la ignorancia ecológica. Lo que se ha intentado desde entonces en muchos laboratorios donde se trabajaba con *E. coli* ha sido desarrollar sistemas muy ingeniosos que hacen uso de los mutantes sensibles a la temperatura. Se trata de obtener mutantes que funcionen normalmente a una baja temperatura de experimentación, por ejemplo 23º C, y que a temperaturas más elevadas, caso de que fueran a parar al intestino humano, se autodestruyan. La mutación que causa esta sensibilidad frente a la temperatura podría formar parte del mismo genoma de *E. coli* o bien ser incorporado a la bacteria a través del portador genético que introduce el nuevo gen.

Las normas de Asilomar contenían principalmente los requisitos que debían cumplir los laboratorios y las precauciones que se debían tomar al trabajar con estos organismos, todo ello en vista a impedir al máximo posibles fugas de material peligroso. Siendo éstas más rigurosas si el material de trabajo son procariotas que intercambian genes con *E. coli*, si se utiliza ADN de mamíferos siendo el receptor *E. coli*, etc. Hay pues, una gradación que depende, como es lógico, del riesgo que implique el experimento.

Los daños que se derivarían del olvido de estas normas o de un descuido desafortunado podrían resultar catastróficos. Supongamos, por ejemplo, que un investigador decidiera favorecer la reducción de precios de los productos alimenticios ahorrando el uso de abono a los cultivadores y empezara a manipular la bacteria *Rhizobium* que fija el nitrógeno del aire. Podría suceder que durante la manipulación además de no conseguir la fijación del nitrógeno (lo que es frecuente en muchas mutaciones) se favoreciera la facilidad de adquirir tumores (característica de algunas bacterias próximas a *Rhizobium*). Si una de estas bacterias saliera de algún modo del laboratorio y resultara ser más resistente (es decir, si hubiese adquirido una ventaja selectiva) que el tipo salvaje de sus especies próximas el resultado podría ser la total destrucción de la cosecha de las plantas en que normalmente se encuentra *Rhizobium*. Se ha mencionado ya el riesgo implícito en los experimentos de "perdigonada". Sabemos que por lo menos algunos de los genes que forman parte del genoma completo de un organismo determinan la síntesis de productos que pueden alterar radicalmente el metabolismo del organismo en que se hayan introducido. Se postula la existencia de virus que causan tumores en los genomas de muchos animales. Si obtuviéramos una bacteria, *E. coli* por ejemplo, con uno de estos genes incorporados en estado no reprimido, ésta se habría convertido en un organismo terriblemente peligroso. Es factible pensar que una de estas bacterias se escapara del laboratorio y se estableciera en una gran población. El desastre sólo sería comparable al de una gran epidemia que no sabemos cómo combatir ya que, a diferencia de lo que ocurre con un producto químico (DDT, PCB), no podemos simplemente parar su síntesis.

En Estados Unidos para concretar y reforzar las normas de Asilomar, el Instituto Nacional de la Salud (N.I.H.) nombró un comité especial de científicos que trabajó arduamente durante un año hasta que en diciembre de 1975 presentó lo que debían ser las normas definitivas de trabajo para aquel país. En conjunto son algo más estrictas que las de Asilomar. Participó la opinión pública en debates abiertos y sus sugerencias fueron consideradas. En junio de este año, el N.I.H. publicó por fin un dossier completo en el que se detallan escrupulosamente las normas sobre el uso de las técnicas que se utilizan para la manipulación de genes en los organismos vivos.

Estas normas afectan a las instituciones donde se lleva a cabo investigación, pero no a la industria, y ejercerán sin duda una gran influencia fuera de sus mismas fronteras. El N.I.H. gracias a los medios económicos que posee puede garantizar el control del cumplimiento de estas normas mediante las ayudas que dispensa a la investigación. El último paso, o sea la aprobación de estas normas por el Congreso, está pendiente de un último informe que deberá presentarse antes de septiembre de 1977.

También han proliferado en Europa los comités para el estudio de las normas sobre las técnicas de manipulación y obtención de ADN recombinante. La EMBO (Organización Europea de Biología Molecular) anunció en enero de este año la creación de un comité para la elaboración de las normas generales de trabajo. Parece, sin embargo, que su función sería la de aconsejar y orientar a los científicos que recurran a él más que el controlar el cumplimiento de unas normas. En Gran Bretaña está a punto de ser aprobado el informe presentado en julio pasado, que fue elaborado por un comité formado por científicos de Universidades y del Consejo Británico.

Mientras tanto, sin embargo, multinacionales como la ICI han montado ya un laboratorio en donde, con finalidades comerciales, se experimenta con ADN recombinante. Las normas británicas a diferencia de lo que ocurre en Estados Unidos serían aplicadas no sólo a centros de investigación, sino también a industrias y, sin duda, ejercerían también una gran influencia en el resto de países europeos. Por todo ello es urgente que estas normas sean aprobadas sin tardanza.

El resto de países comprometidos en esta línea de trabajo deben adoptar antes de que sea demasiado tarde una legislación propia que afecte a todo tipo de organismos.

La sociedad tiene el derecho a ser informada y consultada. Pensemos en el desastre que podría derivarse del empleo de estos organismos en caso de guerra o agresión y, por tanto, su uso debería ser legislado a nivel internacional. Con la "ingeniería genética" los biólogos, al igual que los físicos con el uso de la energía nuclear, han traspasado ya la era de la inocencia. Y, como dijo Oppenheimer, quizá hayan ya caído en el pecado. (R. G.)

Erupciones de rayos gamma cósmicos

Conocidas desde hace sólo cuatro años, se les detecta a un ritmo aproximado de una al mes con dispositivos instalados en los satélites. "Tapan" las demás fuentes de radiación de idéntico tipo; su origen sigue siendo un misterio

Ian B. Strong y Ray W. Klebesadel

La primera descripción del fenómeno que hoy se conoce como erupción de rayos gamma cósmicos se publicó hace más de tres años. Pese a que en la actualidad se ha llegado a acumular un conjunto nutrido de datos de observación, a partir de más de 50 acontecimientos individuales, no existe todavía ninguna explicación, que sea generalmente aceptada, acerca de cómo se generan las erupciones de rayos gamma o en qué parte del universo se originan. Sabemos que trascienden el sistema solar, pero nada hay que pruebe a ciencia cierta que proceden de nuestra propia galaxia o fuera de ella.

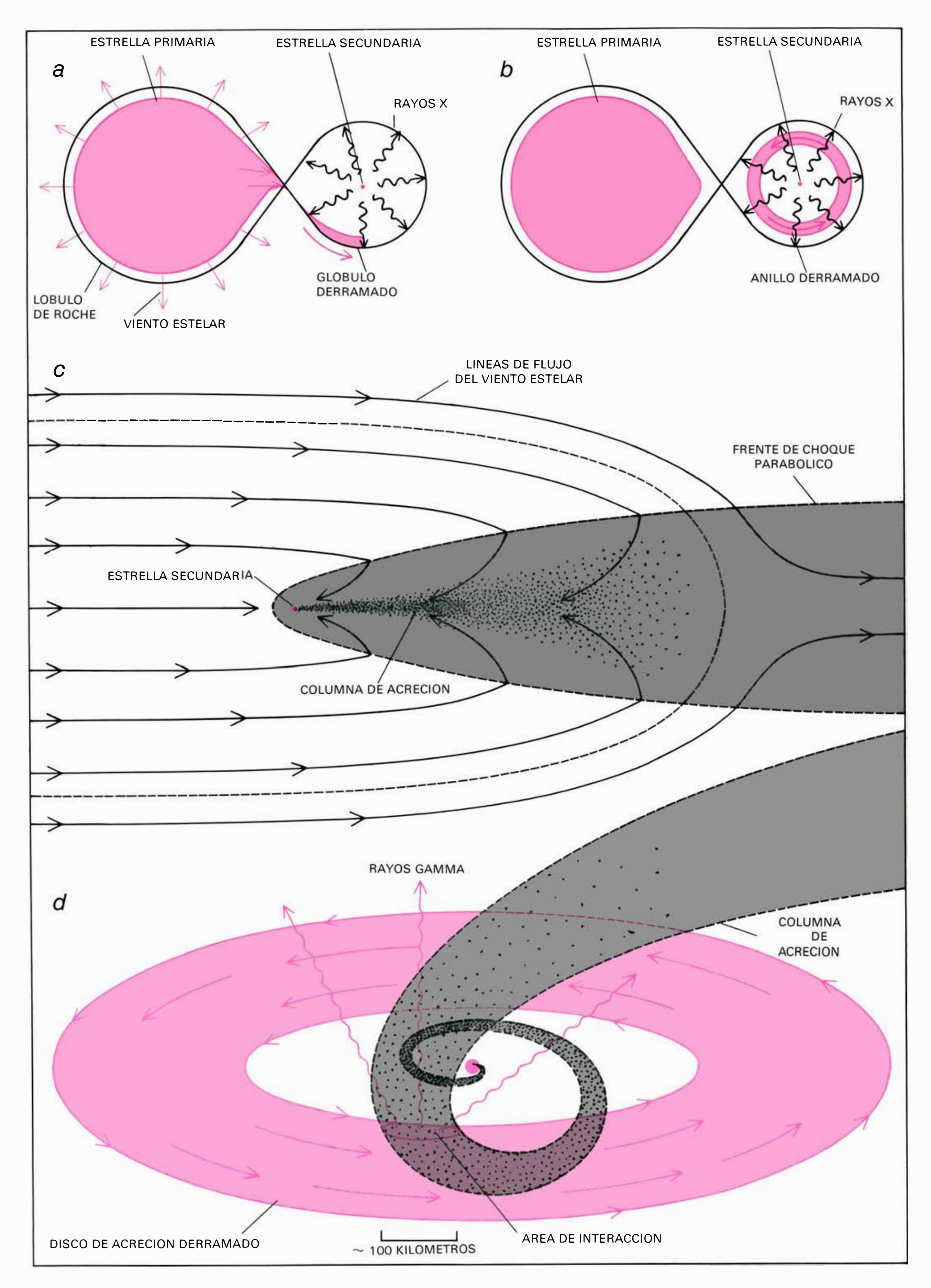
Si las erupciones de rayos gamma cósmicos fueran un efecto sutil, cuya detección requiriera instrumentos muy sensibles, si su duración fuera tan breve que se las pudiese confundir con erupciones de ruido electrónico, o aunque fueran sólo muy poco frecuentes, la persistencia del misterio de su origen sería menos sorprendente. Pero no es éste el caso. Las erupciones de rayos gamma cósmicos son tan intensas que dominan transitoriamente, por varios órdenes de magnitud, a todas las demás fuentes celestes de rayos gamma. Muchas de las erupciones duran más de 10 segundos, lo cual las hace fácilmente visibles, incluso a partir de datos con baja resolución temporal, suministrados por detectores que acumulan fotones durante fracciones grandes de segundo. De hecho salvo unas pocas, todas las erupciones que se conocen fueron detectadas por instrumentos que en principio no estaban destinados a ello. Los sistemas de detección más recientes registran, por término medio, una docena de erupciones al año.

Si cualquier fenómeno comparable hubiera caído en la región visible del espectro, habría sido conocido ya por el hombre primitivo y estaría vívidamente descrito en las obras astronómicas más antiguas.

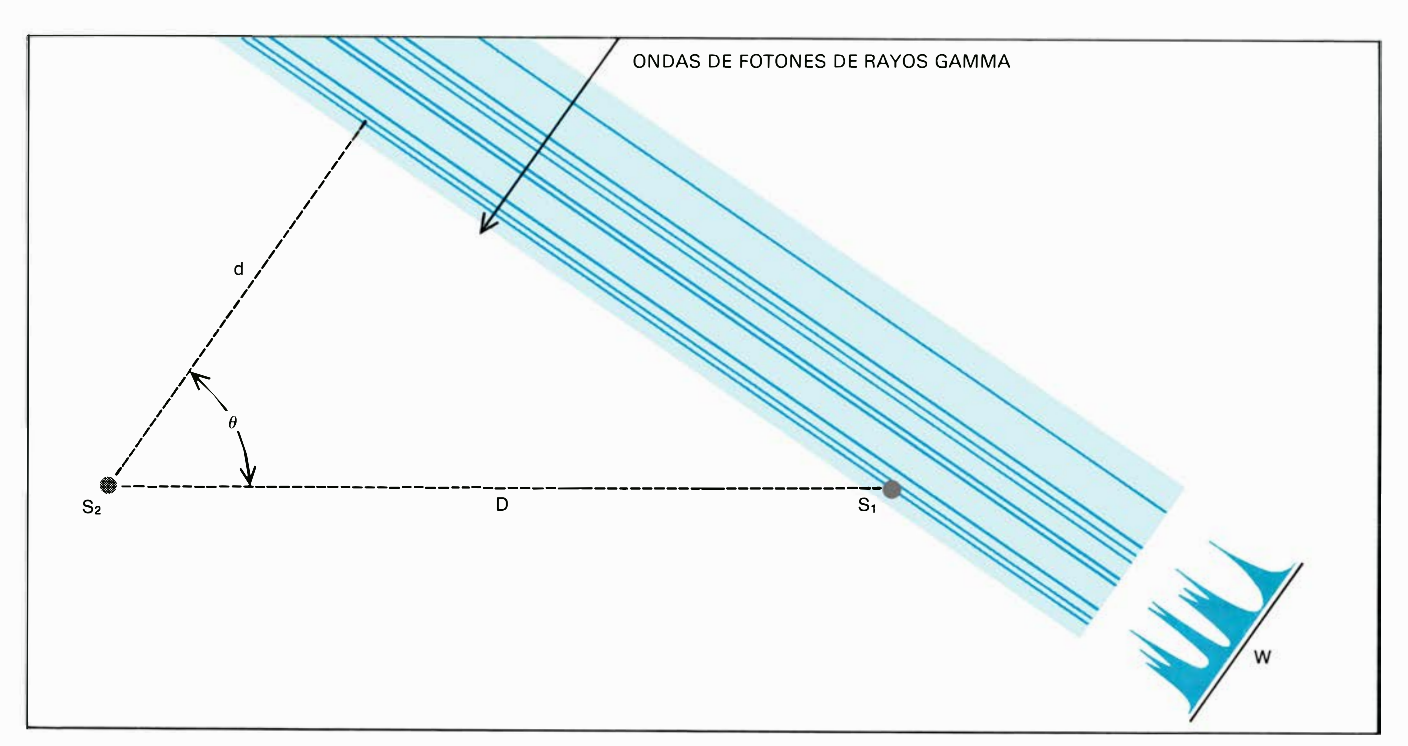
Este descubrimiento tan espectacular e inesperado ha provocado espontáneamente, entre los astrofísicos, una serie de especulaciones bastante incontroladas. Según se clasifiquen las diferentes hipótesis, podría decirse que, hasta el presente, se han propuesto por lo menos 30 modelos distintos, encaminados cada uno de ellos a explicar de dónde proceden las erupciones de rayos gamma y cómo se generan. Como tendremos ocasión de ver, los datos son sumamente apropiados para liberar la imaginación y totalmente inadecuados para ponerle freno.

La historia empezó hace 13 años, cuando los Estados Unidos lanzaron el primer satélite de la serie de los Vela, con el propósito de controlar el cumplimiento del tratado de prohibición de pruebas nucleares de 1963, que impedía a sus firmantes hacer estallar ingenios nucleares tanto en la atmósfera como en el espacio exterior. Los satélites, cuyo instrumental fue diseñado en el Laboratorio Científico de los Alamos, estaban equipados para detectar radiaciones del tipo de las liberadas en las explosiones nucleares; rayos gamma emitidos en el momento de la fisión o de la fusión, radiación térmica (la mayor parte de ella en la región de rayos X del espectro) y neutrones rápidos. Los satélites Vela están emparejados; dos vehículos idénticos circundan la tierra, situados en extremos opuestos de una órbita circular

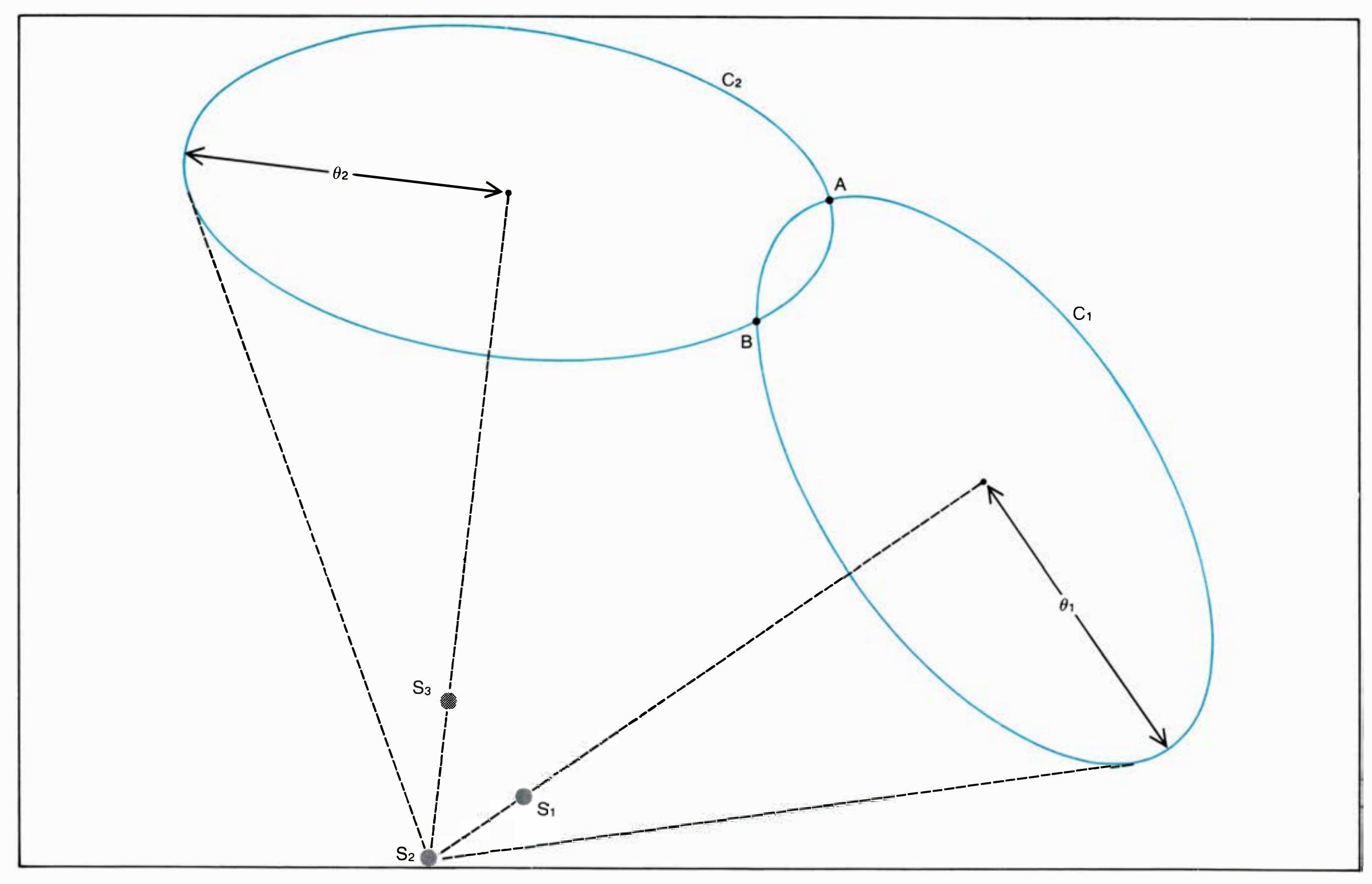
MODELO DE INTERACCION DE PLASMA, propuesto por los autores como una de las fuentes imaginables de los fotones energéticos detectados junto a la tierra como erupciones de rayos gamma. El modelo viene sugerido por la posibilidad de que la fuente de rayos X, conocida como Cygnus X-1 dé también lugar a algunas erupciones. Cygnus X-1 es, con toda seguridad, un sistema doble que consiste en una estrella primaria de grandes dimensiones y en otra secundaria compacta, que puede ser una estrella de neutrones o, como creen algunos, un agujero negro. Un agujero negro es el hipotético objeto que se crearía si una estrella ya agotada colapsara hasta una densidad infinita, produciendo con ello un campo gravitatorio tan fuerte que nada podría encajar, ni siquiera la luz. Sea o no la secundaria un agujero negro, un "viento" estelar de partículas cargadas sopla a partir de la estrella primaria, envolviendo la estrella compacta secundaria *(a)*. Una porción del viento estelar forma una columna de acreción detrás de la secundaria y se enrosca continuamente hacia dentro, en pos de ésta *(c)*. El débito normal de rayos X de Gygnus X-1 procede de la deceleración de las partículas del plasma que se van transfiriendo, en ese sentido, de la primaria a la secundaria. Sin embargo, periódicamente se puede verter glóbulos de plasma de la estrella primaria fuera del lóbulo o límite de Roche, los cuales podrían ser capturados entonces, directamente, por el campo gravitatorio de la secundaria. Los glóbulos forman primero un anillo delgado, que se convierte en un disco fino con un agujero, el cual se contrae lentamente *(b)*. Aunque el disco está definidamente en el plano de la órbita del sistema doble, el plano de la columna de acreción puede, debido a fluctuaciones del viento estelar, desplazarse algo por encima o por debajo de aquél. Cuando ambos planos no coincidan, el plasma derramado del disco en contracción podrá chocar con violencia con el que fluye en la columna de acreción, liberando millones de electrón-voltios de energía por cada partícula en colisión *(d)*. Así resulta una erupción de radiación gamma, tal vez aumentada por los fotones de alta energía procedentes de la desintegración de los piones formados en colisiones protón-protón. La producción brusca de radiación tiende a separar los dos plasmas, interrumpiendo consiguientemente la liberación de ésta. Ello podría dar cuenta de la estructura fina que se ha visto en las erupciones de rayos gamma. La erupción de rayos gamma ocurre sólo cuando las intensidades de las dos corrientes de plasma son comparables. Plausiblemente, el disco derramado dominará, y la radiación X usual alterará sus características hasta que se haya agotado la reserva de plasma (puede tardar varios meses). Tales cambios se han observado en el débito de rayos X de Cygnus X-1, y algunos pueden coincidir con erupciones de rayos gamma.



a
ESTRELLA PRIMARIA
ESTRELLA SECUNDARIA
RAYOS X
GLOBULO DERRAMADO
LOBULO DE ROCHE
VIENTO ESTELAR
b
ESTRELLA PRIMARIA
ESTRELLA SECUNDARIA
RAYOS X
ANILLO DERRAMADO
c
LINEAS DE FLUJO DEL VIENTO ESTELAR
FRENTE DE CHOQUE PARABOLICO
ESTRELLA SECUNDARIA
COLUMNA DE ACRECION
d
RAYOS GAMMA
COLUMNA DE ACRECION
DISCO DE ACRECION DERRAMADO
~ 100 KILOMETROS
AREA DE INTERACCION



EL ANGULO DIRECCIONAL de un suceso de rayos gamma puede determinarse si los tiempos de llegada de los fotones de la erupción han sido medidos con precisión por dos satélites cuyas posiciones en el espacio y cuya separación, *D*, se conozcan también de un modo preciso. La diferencia entre los tiempos de llegada da la distancia *d*. Conociendo *d* y *D*, se puede obtener θ, el ángulo que forma la onda plana incidente con una recta que pase por los dos satélites. Debe observarse, sin embargo, que el ángulo θ puede hacerse girar 360 grados en torno al segmento *D*. De ello resulta que los datos de dos satélites sólo pueden especificar una circunferencia sobre la esfera celeste. Si el suceso de rayos gamma tiene una estructura temporal (*W*), pueden usarse los rasgos comunes para precisar los tiempos de llegada de ambos satélites.



UNA NUEVA RESTRICCION EN LA LOCALIZACION de una fuente puede obtenerse si un suceso queda registrado por tres satélites. La diferencia entre los tiempos de llegada a un par de satélites, S_1 y S_2, por ejemplo, da el ángulo θ_1, que determina el tamaño de una circunferencia en el espacio C_1, cuyo centro está sobre la recta que une ambos satélites. La diferencia de tiempos de llegada a otro par de satélites, S_2 y S_3, da el ángulo θ_2, que determina el tamaño de un segundo círculo, θ_2, cuyo centro está sobre la recta que une el segundo par de satélites. Puesto que la fuente debe estar sobre las referidas circunferencias, debe encontrarse en uno de los puntos de intersección, el *A* o el *B*. Sin embargo, se puede eliminar uno de estos puntos de intersección si el suceso es registrado por un cuarto satélite que no esté en el mismo plano que S_1, S_2 y S_3, o si se puede ver que uno de los dos puntos estaba oculto para otro satélite que también haya detectado la erupción.

de 250.000 kilómetros de diámetro, de modo que ninguna región del espacio queda oculta por nuestro planeta. Durante una porción muy grande de los últimos 13 años, el sistema Vela ha dispuesto de varios satélites, lo cual, al proporcionar redundancia, asegura que la radiación detectada no sea, simplemente, algo originado en las proximidades de cualquier satélite.

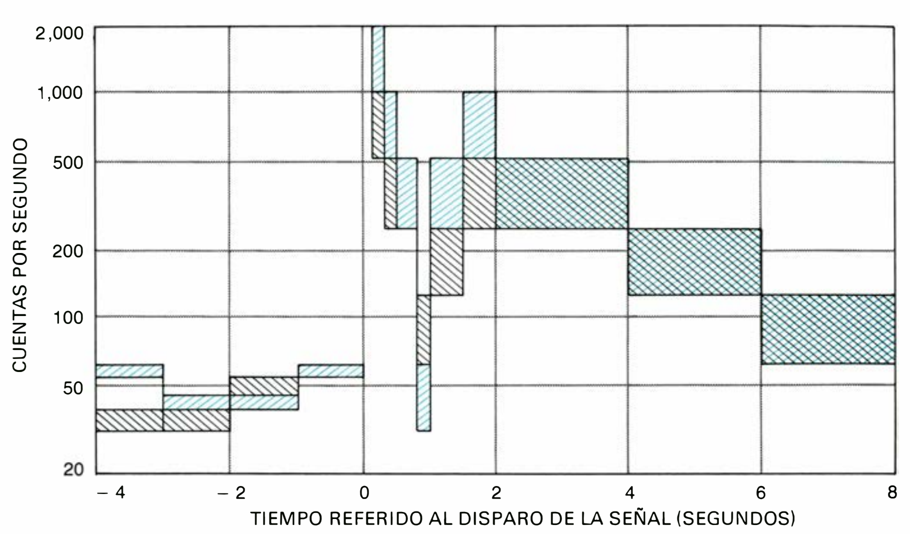
Los satélites Vela posteriores se mejorarán con la adición de nuevos detectores y con un superior tratamiento de las señales, que permitió hacer un estudio más detallado de la radiación natural del entorno, por sus posibles efectos sobre el instrumental y también por la posibilidad de que la naturaleza imitara el tipo de señales producido por las armas nucleares. Los nuevos datos iban a ser también del mayor interés científico.

Cómo se detectaron las erupciones

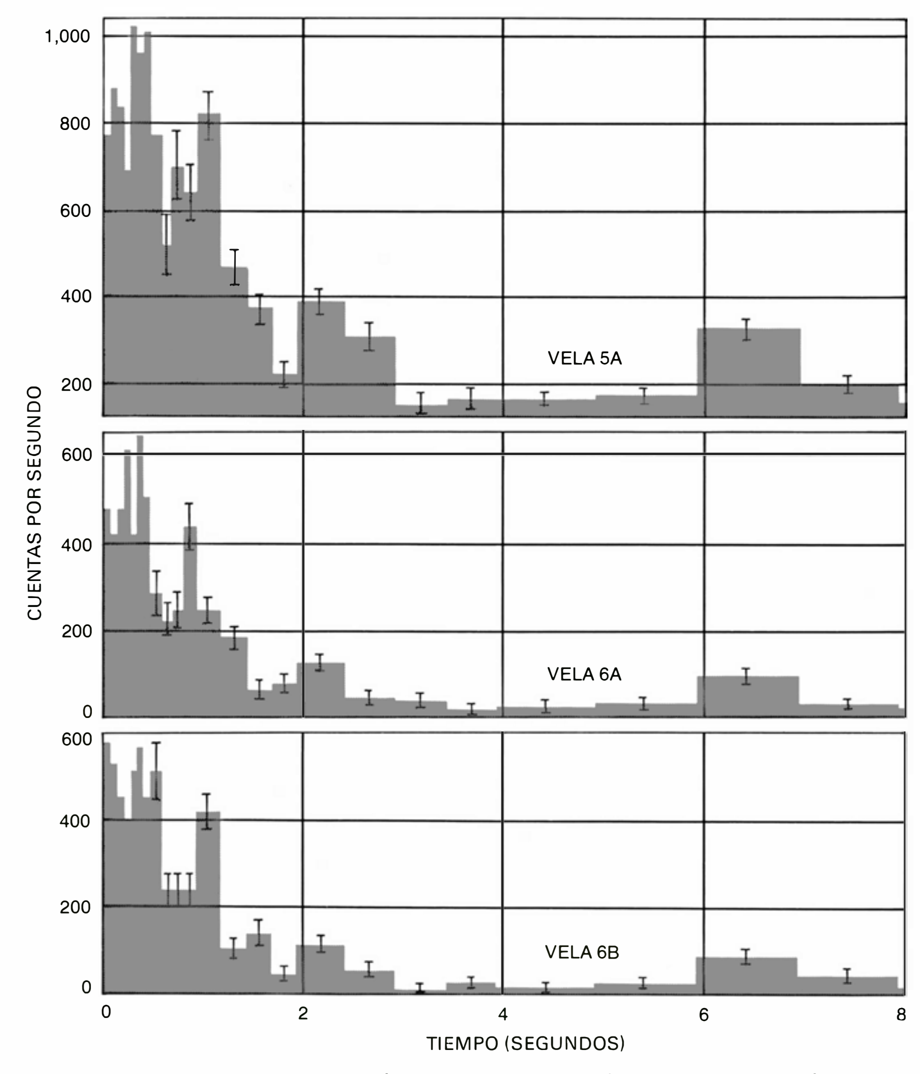
Todos los satélites Vela dieron un resultado excepcional, con una perduración en su esperanza de vida mayor de la asignada (basada en la vida media de otros satélites de la misma época). Sin embargo, este resultado tuvo la desventaja de generar muchos más datos de lo que se había esperado, con lo cual el tratamiento de los rutinarios se fue retrasando mucho. Ese postergamiento sufrieron los datos suministrados por un sistema de detección de rayos gamma, incorporado a la cuarta serie de los satélites Vela. El sistema se diseñó para estudiar los detalles de erupciones de rayos gamma que duraran varios segundos, minutos incluso, en el caso de que tales erupciones existieran.

En 1969 se registraba la primera erupción de est tipo, perdida entre los datos pertenecientes a 1967. Los dos satélites Vela 4 detectaron señales más o menos idénticas, lo cual indicaba que la fuente estaba aproximadamente a la misma distancia de cada uno de ellos. Pudiera ser que estuviera cerca de la tierra, pero igualmente que se encontrara junto al sol. Conociendo con suficiente precisión los tiempos de llegada de la señal a cada uno de los satélites, esto hubiera podido comprobarse, puesto que ambos satélites están siempre a la misma distancia de la tierra, pero suelen diferir en su distancia respecto del sol.

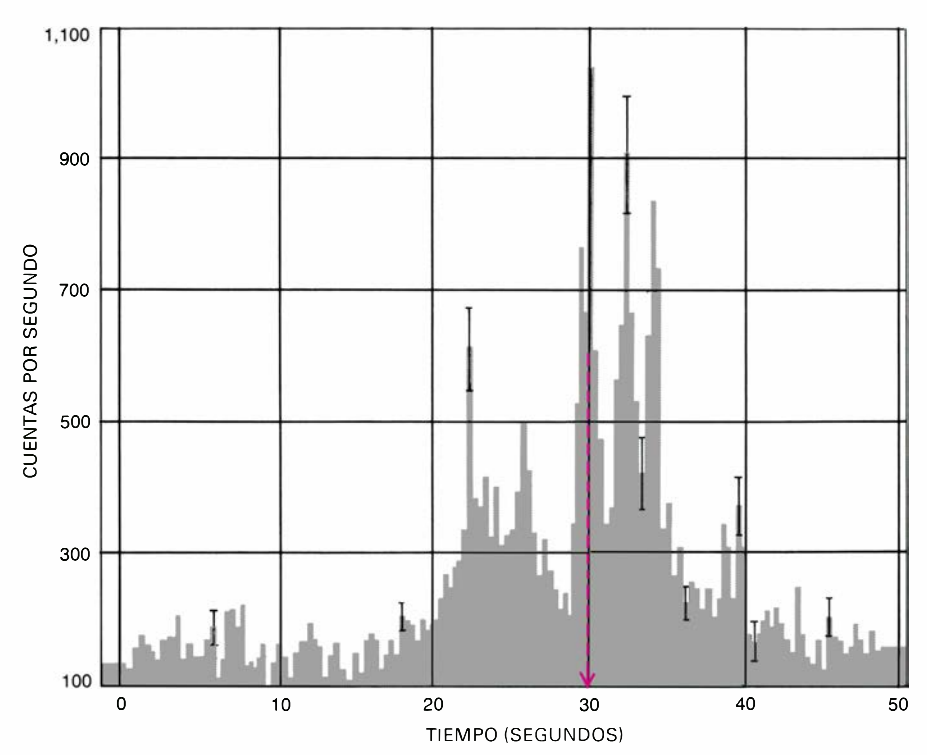
Los satélites posteriores se modificaron en el sentido de que pudieran registrar la época de tales acontecimientos con una precisión de 0,016 segundos. Hacia 1972, una búsqueda cuidadosa de acontecimientos fortuitos, que hubieran tenido lugar casi simultáneamente en los



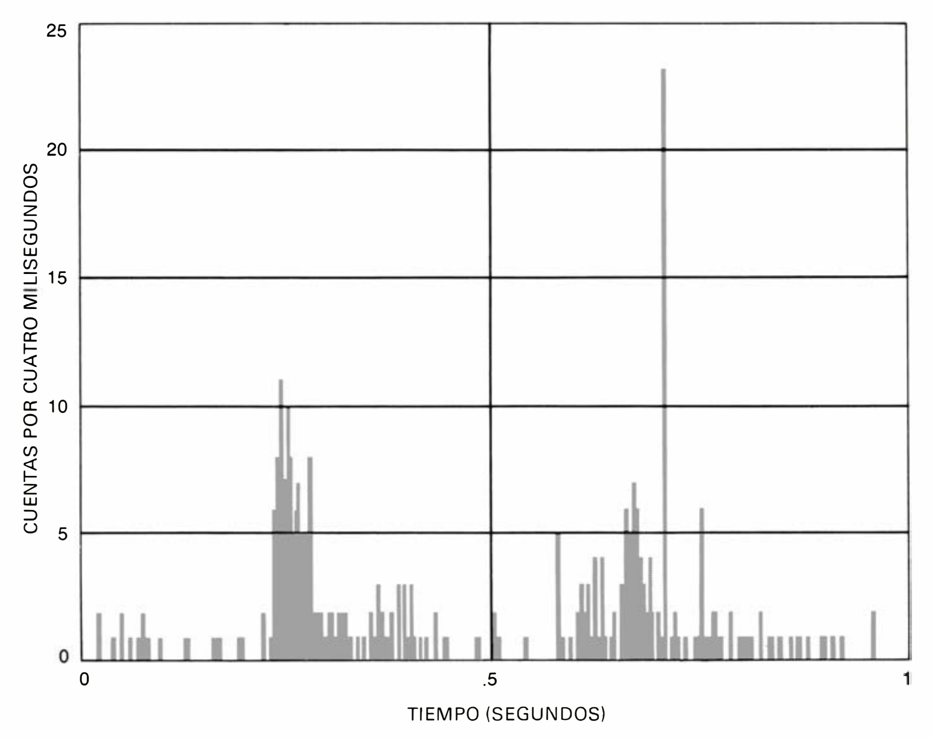
LA PRIMERA ERUPCION DE RAYOS GAMMA se observó el 2 de julio de 1967, con dos satélites Vela que llevaban instrumentos destinados a detectar radiaciones del tipo de las que produciría una explosión nuclear en el espacio. Respondiendo casi simultáneamente, los detectores embarcados en los satélites, que se movían en órbita a unos 250.000 kilómetros el uno del otro, dieron lugar a respuestas casi idénticas. Los detectores responden a fotones que estén dentro del intervalo de energías que va de 0,2 MeV (millones de electrón-voltios) a 1,5 MeV.



TRIPLE REGISTRO de una erupción de rayos gamma, efectuado por tres satélites Vela el 22 de agosto de 1970. Los detectores de rayos gamma de yoduro de cesio sólo empiezan a almacenar datos si el ritmo de cuenta sube bruscamente durante algún cuarto de segundo. Las cuentas se almacenan entonces a lo largo de una sucesión de intervalos de tiempo que se van doblando sistemáticamente en longitud. Puede verse que la estructura fina de la erupción, que fue la séptima en observarse, es registrada de la misma manera por los detectores de los tres satélites.



UN REGISTRO FORTUITO de una erupción de rayos gamma, que apenas si disparó los detectores Vela, cursaron los dispositivos embarcados en el *Apolo 16*, el 27 de abril de 1972, justo antes de que el vehículo espacial tripulado entrara en la atmósfera de la tierra, tras su viaje lunar. Los detectores del *Apolo 16*, mayores y más sensibles que los de los satélites Vela, suministraron la imagen más detallada disponible de una erupción de rayos gamma, previa al primer suceso registrado por el satélite alemán occidental *Helios-2 (véase la ilustración de abajo)*. Tras alcanzar un pico de más de 1000 fotones por segundo, la cuenta cayó a cero durante un período de 18 milisegundos *(línea discontinua en color)* y luego volvió casi al nivel anterior; el breve período a cero (18 milisegundos) no puede representarse de manera adecuada a esta escala.



LA PRIMERA ERUPCION DE RAYOS GAMMA DE 1976, detectada el 23 de enero, fue también la primera en ser registrada por un instrumento específicamente construido para este fin. El instrumento fue diseñado por el Goddard Space Flight Center y embarcado en el *Helios-2*. El suceso registrado por el *Helios-2* es típico por lo que se refiere a su estructura general y por la tendencia de las espículas finas a llegar en grupos. El detector cuenta rayos gamma por períodos consecutivos, de cuatro milisegundos cada uno. Aquí sólo mostramos el primer segundo de la erupción, que duró varios segundos. El pico más alto consta de 23 cuentas, con sólo una cuenta en los dos intervalos de tiempo adyacentes. Por lo tanto, la señal del pico probablemente no duró más de dos milisegundos, incluyendo la subida y la bajada. Ello implica que la fuente es notablemente pequeña, de 300 kilómetros de diámetro, a lo sumo.

datos de dos, tres o cuatro satélites, puso de manifiesto un número sorprendente de ellos, que no sólo eran casi contemporáneos sino que también eran del mismo orden de intensidad y tenían la misma distribución de intensidades en función del tiempo. Si las fuentes eran exteriores a la órbita de los satélites, tenían que estar muy lejos; de otro modo, el detector más próximo hubiera registrado una señal más fuerte que los más alejados. Además, la cronología de acontecimientos en dos o más satélites suministró información direccional, que mostraba que los rayos gamma no podían venir ni de la tierra, ni de la luna o del sol, ni tan siquiera de ninguno de los planetas.

Se adquirió entonces la certeza de que se trataba de un fenómeno no observado hasta el momento, originado probablemente fuera del sistema solar y que implicaba evidentemente enormes cantidades de energía en la fuente. La primera reseña de este desconcertante fenómeno se publicó en 1973. Se tiene información de un total de 55 erupciones, no sólo a partir de los satélites Vela, sino también a partir de otros muchos satélites y globos.

Dirección de las fuentes

Los detectores de rayos gamma embarcados en los satélites Vela consisten en seis cristales de yoduro de cesio, que emiten luz visible al ser penetrados por los rayos gamma. La luz es convertida en pulsos electrónicos y registrada. Los seis cristales, que trabajan combinados, son igualmente sensibles a la radiación procedente de cualquier dirección. Debido a que se podía esperar que la intensidad de la radiación estudiada decrec era más o menos exponencialmente con el tiempo, los detectores se diseñaron para que almacenaran sus cuentas en una sucesión correspondiente a intervalos de tiempo cada vez más largos, empezando por 1/64 segundos y continuando hasta un total de 15 minutos. Los detectores sólo empiezan a almacenar datos cuando el ritmo de cuentas aumenta en gran medida durante un determinado cuarto de segundo.

Es evidente que si dos satélites detectan simultáneamente un pulso de rayos gamma, la fuente de la radiación debe equidistar de ambos. En efecto: ha de estar situada sobre un plano perpendicular, en su punto medio, al segmento que une a los dos satélites. Si los tiempos de llegada no son simultáneos, se puede determinar fácilmente el ángulo que forma la detección de la onda incidente con el referido segmento. Este ángulo,

proyectado sobre la esfera celeste, determina una circunferencia, sobre la cual debe estar situada la fuente.

Si la radiación gamma viene detectada por un tercer satélite, el ángulo entre la dirección de la onda y la recta que una al tercer satélite con uno cualquiera de los otros dos determina el tamaño y la posición de una segunda circunferencia, que corta a la anterior. La fuente estará en uno de los dos puntos de intersección de ambas. No se puede saber en cuál de ellos, a menos que un cuarto satélite, que no esté situado en el mismo plano que los otros tres, registre también el suceso. A veces se dispone de otro tipo de información, para decidir entre las dos posibilidades. Varios satélites (pero no los Vela) pasan tan cerca de la tierra que sólo les queda visible una mitad del cielo en cada momento. Si uno de estos satélites dispone de un detector que responda de algún modo a la erupción de rayos gamma, y en ese momento sólo es visible una de las dos posibles fuentes, se puede saber la dirección correcta. En base a los datos del detector de rayos X "duros" del satélite OSO-6, G. G. C. Palumbo, G. Pizzichini y sus colegas, de la Universidad de Bolonia, fueron capaces de eliminar una de las dos posibles fuentes, para un cierto número de erupciones de rayos gamma. Otras elecciones se hicieron en base a las respuestas de los satélites *Uhuru,* SAS-2 y OGO-5.

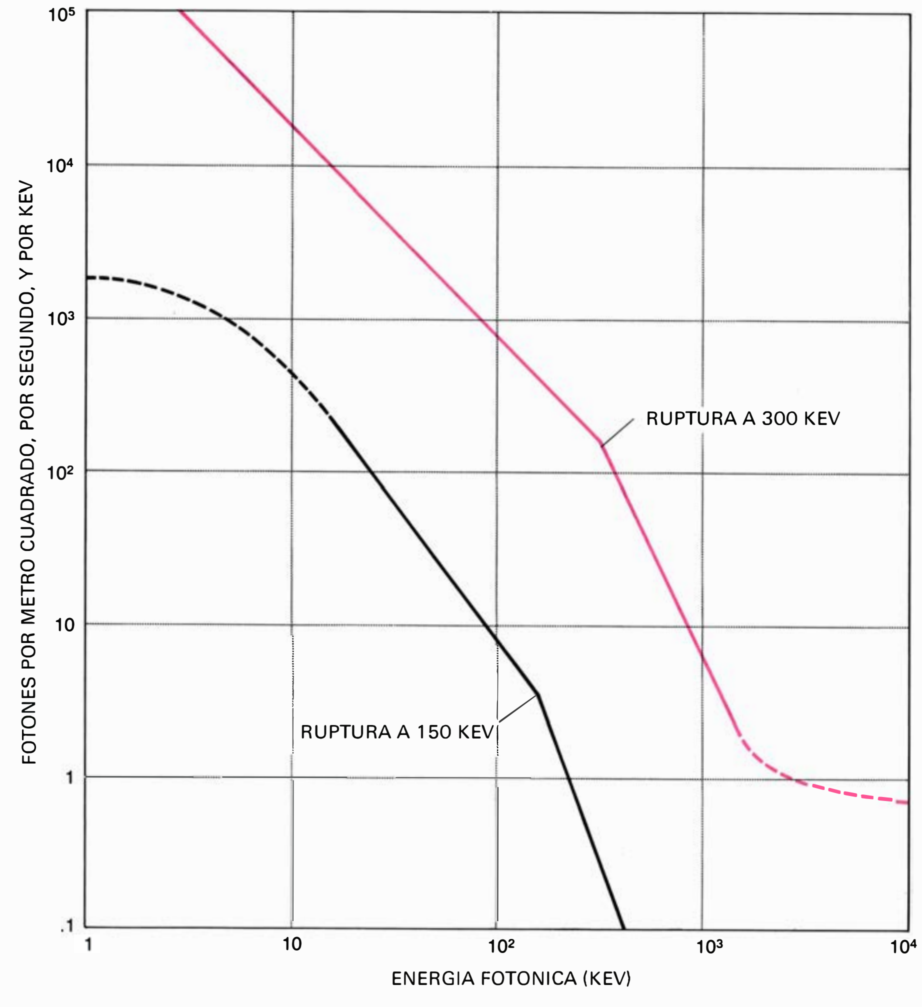
La dirección de una de las fuentes de erupción de rayos gamma quedó confirmada, independientemente, por un telescopio de rayos X con sensibilidad direccional, embarcado en el satélite OSO-7. Hay otro acontecimiento que es de particular interés, pues, pese a ser tan débil que apenas llegó a disparar los detectores Vela, fue registrado con gran detalle por los grandes dispositivos, dotados de gran sensibilidad de la nave espacial tripulada *Apolo 16,* cuando ésta volvía de la luna, justamente antes de entrar en la atmósfera terrestre, el 27 de abril de 1972.

Naturaleza de las erupciones

Las erupciones presentan una gran variedad. El tiempo característico de aumento del pulso inicial varía de 0,002 segundos a más de 1 segundo. El pulso inicial dura desde menos de 0,01 segundos hasta 10 segundos. Puede haber un solo pulso o hasta cinco, y más, claramente separados; la erupción entera puede durar entre 0,1 segundos y más de 100. A medida que se ha ido disponiendo de mejores instrumentos, con mayor resolución en tiempo, se ha visto que la estructura de las erupciones es incluso más compleja que lo que se había creído en un principio. La cantidad de energía que llega a los satélites oscila desde 3×10^{-6} ergios por centímetro cuadrado y por segundo hasta 300 veces más, con un valor medio de aproximadamente 10^{-4} ergios por centímetro cuadrado y por segundo.

Para comprender cualquier proceso astrofísico en el que intervenga la radiación, es esencial disponer de buena información sobre la distribución por longitudes de onda de la energía emitida. Y ello tanto si se trata de la región visible del espectro como si corresponde a la de rayos X y gamma. Los detectores Vela proporcionan sólo un mínimo de información espectral. Los detectores Vela 5 responden a fotones de energías comprendidas entre 150 y 750 kiloelectrón-voltios (keV), mientras que los Vela 6 lo hacen a fotones con energías de 300 a 1500 keV. Los dos intervalos, pues, se solapan. La situación es análoga a cuando se usan placas fotográficas, sensibles las unas al rojo, las otras al azul, para estudiar estrellas en la región visible. A partir de la diferencia en brillo de una misma estrella, según se use una u otra placa, podremos decir en qué medida la estrella es más rojiza o más azulada, pero no mucho más. De modo análogo, comparando el número de fotones contados durante una misma erupcion por un detector Vela 5 con los contados por un Vela 6, podremos decir si la radiación es "blanda" o "dura" (de más baja o más alta energía, respectivamente).

Los satélites IMP 6 e IMP 7, que miden el brillo en ocho o nueve intervalos de energía diferentes, han suministrado espectros de la intensidad, referida a la energía, para cerca de 30 de las más lar-



EL ESPECTRO DE UNA ERUPCION DE RAYOS GAMMA *(en color)* **tiene un parecido con el espectro medio** *(en negro)* **de los rayos X emitidos por Gygnus X-1, la fuente de rayos X de la que se sospecha que contiene un agujero negro. El espectro de rayos gamma está basado en las mediciones efectuadas por los instrumentos embarcados en el** *Apolo 16.* **Ambas curvas están trazadas en fotones por metro cuadrado, por separado y por KeV (kiloelectrón-voltios), lo cual subraya el extremo de baja energía del espectro. El cambio de pendiente del espectro de rayos gamma, a 300 KeV, es evidentemente real. Puede resultar interesante que el espectro de Cygnus X-1 presente una ruptura similar, pero a más baja energía. Las regiones de trazos de ambas curvas mostradas representan dudas en los datos de observación o en la interpretación de éstos.**

gas erupciones de rayos gamma. Thomas L. Cline y U. D. Desai, del Goddard Space Flight Center, han llegado a la conclusión, a partir del estudio de esos datos, de que los espectros, una vez promediados sobre toda la erupción, son todos muy parecidos. La medición espectral única más detallada, la registrada por el *Apolo 16,* fue cuidadosamente analizada por Albert E. Metzger y sus colaboradores, en el Jet Propulsion Laboratory del Instituto de Tecnología de California. Aunque la curva no presenta ninguna prueba de rayos de emisión, tampoco es completamente regular. Por término medio, los fotones son de energías mucho mayores que las típicas de las fuentes cósmicas de rayos X y el número de fotones decrece menos abruptamente al aumentar la energía. Se insinúa incluso un cierto número adicional de fotones a energías de algunos millones de electrón-voltios, muy por encima de lo que podría esperarse viendo sólo la parte principal de la curva.

La erupción del *Apolo 16* aparenta una estructura temporal más complicada que la de otros sucesos, aunque puede que ello se deba más a la alta calidad de los datos que a la naturaleza de la erupción. La radiación puede arrancar y parar bastante bruscamente, en cuestión de milisegundos. Dentro del pico más alto de los mejores datos del *Apolo 16,* la radiación cesó completamente durante 18 milisegundos. Un corolario importante de la existencia de fluctuaciones tan rápidas es que la región que emite los rayos gamma debe ser pequeña. Si la cuenta de fotones sube o baja bruscamente en *t* segundos, se puede deducir que el diámetro efectivo de la fuente no excede la distancia atravesada por la radiación en ese mismo intervalo. En base a los datos del *Apolo 16,* puede afirmarse que el diámetro máximo de la fuente era de 1500 kilómetros, poco más de una décima parte del diámetro de la tierra.

No se puede afirmar si todas las fuentes de las erupciones de rayos gamma son o no de tamaños parecidos. En cualquier caso, la fuente de la erupción registrada el 23 de enero de 1976 por el detector embarcado en el *Helios-2* establece un nuevo récord de pequeñez. Los fotones de los rayos gamma se contaron por períodos consecutivos de 4 milisegundos. Durante el referido suceso, el detector registró, en promedio, un fotón por período de cuenta. En uno de los períodos se contaron 23, con un fotón en el período anterior y ninguno en el siguiente. Es sumamente improbable que esa breve "punta" de rayos gamma durara los 4 milisegundos enteros. Se hubiera repartido, con toda verosimilitud, entre dos períodos de cuenta. Su duración fue, probablemente, de 2 milisegundos, o menos, incluyendo el crecimiento, la caída y cualquier emisión sostenida que pudiera tener lugar durante el máximo. Llegamos así a la conclusión de que el tiempo de crecimiento fue de un milisegundo, o menos, en cuyo caso el diámetro máximo de la fuente sería sólo de 300 kilómetros.

	UNIDADES X	INTERVALO	VALOR TIPICO
ENERGIA TOTAL EN LA TIERRA	ERGIOS POR CENTIMETRO CUADRADO	DE 3×10^{-6} A $1,5 \times 10^{-3}$	10^{-4}
NUMERO DE PULSOS DISTINTOS		DE 1 A MAS DE 5	2
DURACION DEL PRIMER PULSO	SEGUNDOS	DE < .01 A 10	DE 3 A 5
DURACION TOTAL DE LA ERUPCION	SEGUNDOS	DE < 0.1 A ~ 100	10
TIEMPO DE SUBIDA DEL PRIMER PULSO	SEGUNDOS	DE .002 A 3	.2
INTENSIDAD MAXIMA	ERGIOS POR CENTIMETRO CUADRADO Y POR SEGUNDO	DE 10^{-4} A 5×10^{-4}	

CONSIDERABLE VARIABILIDAD presentan las cincuenta erupciones, aproximadamente, registradas desde el primer suceso en 1967. La tabla muestra intervalos y valores típicos de los principales parámetros. Las cifras que se dan para el flujo total de energía en la tierra lo son suponiendo que la medición se haga fuera de la atmósfera terrestre y que el espectro de energías, sea parecido al dado en la página anterior. La estrechez del intervalo de máximas de energía, que varían sólo en un factor de aproximadamente cinco, se debe, probablemente, a la reducida capacidad que poseen los detectores actuales para poder registrar las erupciones más débiles.

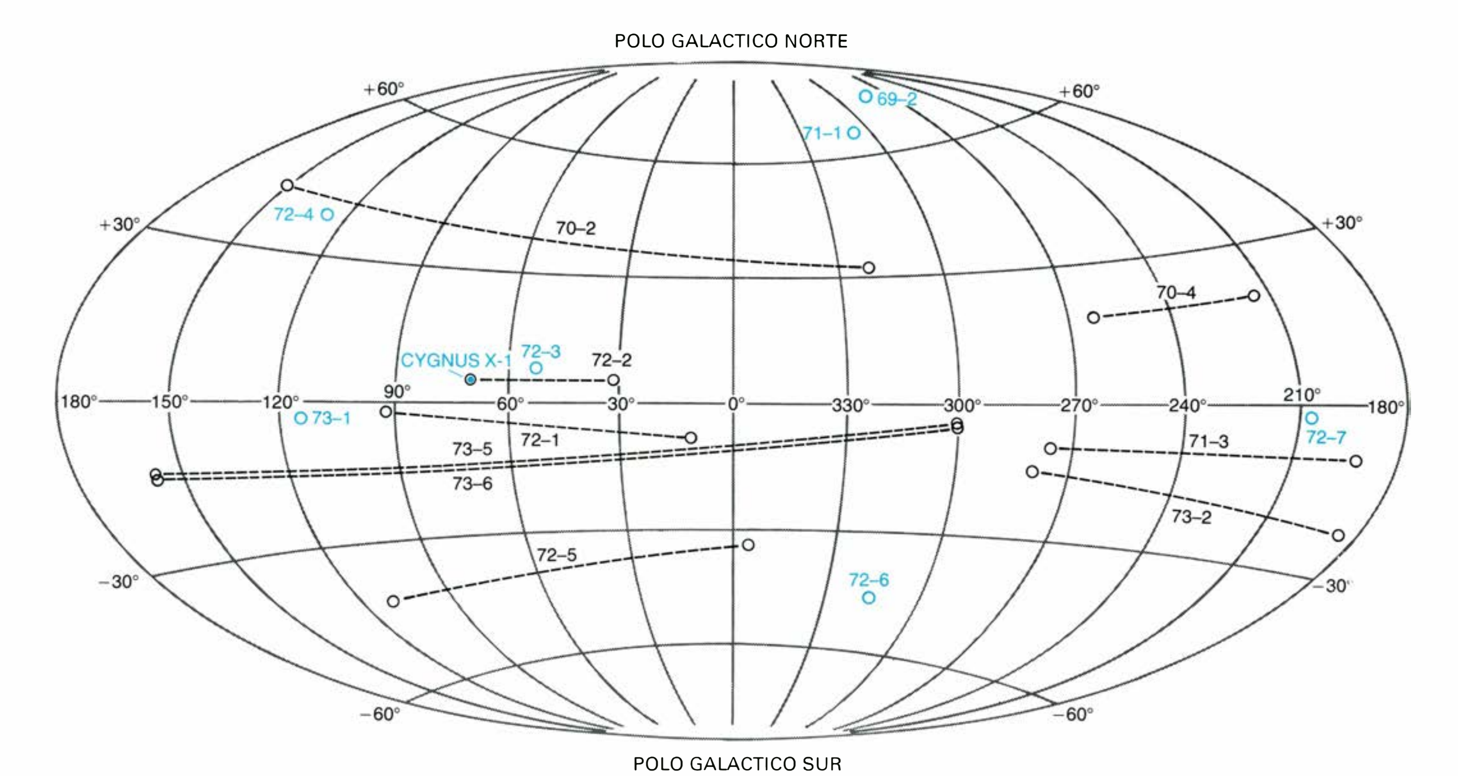
DISTANCIA (PARSECS)	ENERGIA POR ERUPCION (ERGIOS POR SEGUNDO)	POSIBLE FUENTE
1	10^{34}	ESTRELLAS MAS PROXIMAS
10	10^{36}	ESTRELLAS DE RAYOS X ERUPTIVAS (MINIMO)
10^2	10^{38}	MAXIMO PARA ESTRELLAS DEL DISCO GALACTICO (ESTRELLAS DE LA POBLACION I)
10^3	10^{40}	ESTRELLAS DEL HALO GALACTICO (ESTRELLAS DE LA POBLACION II)
10^4	10^{42}	CUMULOS GLOBULARES
10^5	10^{44}	OBJETOS EXTRAGALACTICOS MAS CERCANOS
10^6	10^{46}	GRUPO LOCAL DE GALAXIAS
10^7	10^{48}	CUMULOS DE GALAXIAS

LA ENERGIA TOTAL EMITIDA en la fuente, en una erupción de rayos gamma, depende de su distancia, que por ahora no ha sido determinada para ninguno de estos desconcertantes sucesos. Suponiendo que el flujo de una erupción de rayos gamma, medido cerca de la tierra, sea de unos 10^4 ergios por centímetro cuadrado y por segundo, y que la fuente irradia energía uniformemente en todas direcciones, el flujo total en la fuente puede variar desde 10^{34} ergios por segundo (si la fuente está a la distancia de una de las estrellas más cercanas) hasta más de 10^{48} ergios por segundo (si la fuente está en alguna parte fuera del Grupo Local de galaxias).

Energías de las fuentes

Es obvio que si conocemos la cantidad de energía que llega al detector seremos capaces de calcular el total emitido por la fuente, siempre que la energía se emita uniformemente en todas las direcciones (cosa de la que no tenemos por qué dudar) y que conozcamos también la distancia de la fuente. Y ahora surge la dificultad. A menos que las fuentes estén muy cerca de los detectores, lo cual no es el caso, no podemos determinar directamente su distancia. Esta se podría establecer sólo si una erupción de rayos gamma viniera inequívocamente asociada con un objeto celeste cuya distancia se hubiera registrado por otros medios. Sin embargo, se puede calcular la energía que debe radiarse, suponiendo que la fuente esté a las distancias típicas de distintas clases de objetos celestes.

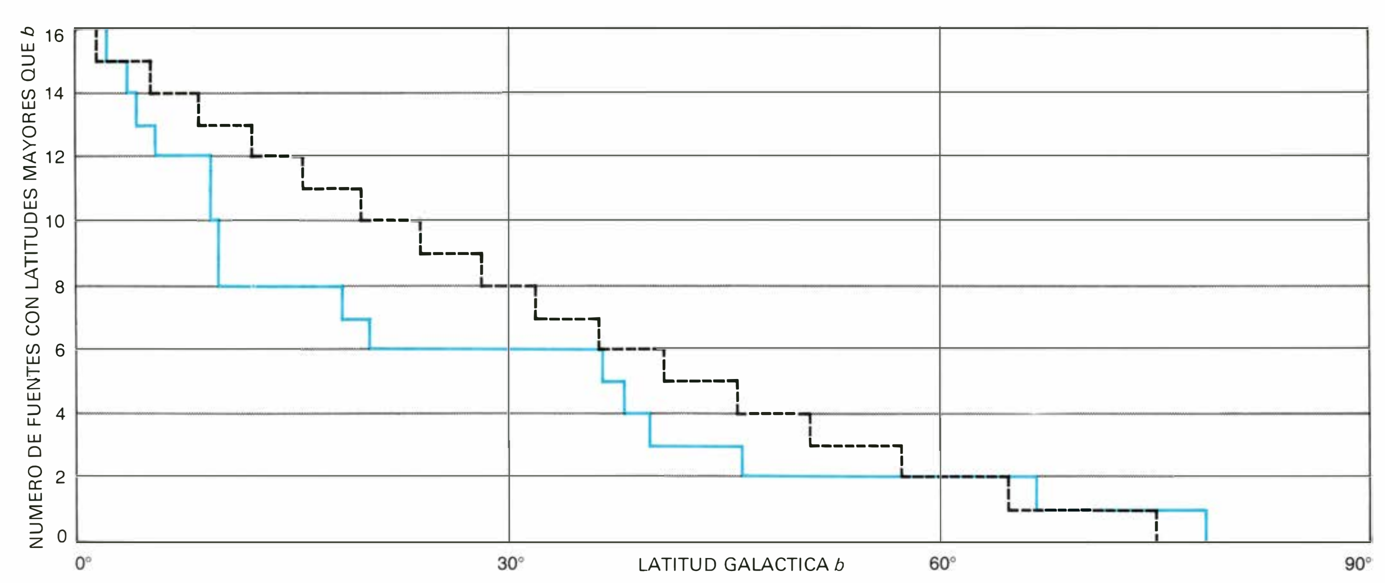
Incluso para distancias moderadas de las fuentes, la energía liberada es enorme. Sobre la base de un flujo típico de rayos gamma de 10^{-4} ergios por centímetro cuadrado y por segundo en el detector del satélite, la energía liberada por una fuente que estuviera situada a la distancia de las estrellas más próximas (algo más de un parsec, o sea 3,26 años-luz) sería 10^{34} ergios por segundo. Si la fuente estuviera a 1000 parsecs, que es sólo una décima parte de la distancia del sol al centro de la galaxia, la energía liberada sería de 10^{40} ergios por segundo. A este ritmo, la energía de los rayos X



LAS POSICIONES DE LAS FUENTES de rayos gamma han sido difíciles de determinar univocamente y con alta precisión. Como se ha explicado en las ilustraciones de la página 46, si una erupción de rayos gamma es detectada por tres satélites, la posición de la fuente puede asignársele a cualquiera de los dos puntos en que se cortan ambas circunferencias. Los círculos en blanco señalan tales posiciones alternativas para nueve sucesos representadas en coordenadas galácticas (Los dos primeros dígitos indican el año del suceso; el tercer dígito designa el orden del descubrimiento, dentro del año). En siete casos, marcados con círculos de color, se conocen las posiciones únicas debido a que ha sido posible eliminar una de las dos alternativas porque la tierra la ocultaba de un cuarto satélite, cuya órbita era muy baja.

y gamma emitidos por la fuente sería unos 2,5 millones de veces mayor que la radiada por el sol en todas las longitudes de onda. Teniendo en cuenta que la fuente es menor que la tierra, debería estar radiando al menos diez mil millones de veces más energía por unidad de superficie que el sol. Si las fuentes están en otras galaxias, como algunos han sugerido, las cifras anteriores deberían incrementarse en seis órdenes de magnitud, por lo menos.

Las erupciones de rayos gamma no presentan correlación alguna en su dirección ni con estrellas próximas ni con ninguna de las estrellas cercanas de las que se sabe tienen características peculiares, de modo que nos vemos constreñidos a emplear argumentos estadísticos para inferir distancias. Un método sería, simplemente, estudiar la distribución de intensidades de las erupciones de rayos gamma registradas hasta ahora. Siempre que las fuentes estén distribuidas uni-



LAS FUENTES DE LAS ERUPCIONES DE RAYOS GAMMA parecen, a primera vista, estar distribuidas bastante al azar por el cielo. Una prueba sencilla sobre la posibilidad de acumulación consiste en ignorar las longitudes, la información sobre las cuales es en todo caso rara, y trazar una curva que represente cuántas erupciones están situadas a más de un cierto número de grados, *b*, contados a partir del ecuador galáctico. La curva escalonada de color muestra la distribución de los 16 sucesos, utilizando latitudes promedio para los nueve sucesos con alternativas en la posición. La curva en negro muestra la distribución que sería de esperar si los sucesos estuvieran isótropamente dispersos, es decir, si no tuvieran dirección preferente en el espacio. Aunque la muestra de posiciones es muy reducida para ser concluyente, la comparación de las dos curvas escalonadas sugiere que las erupciones de rayos gamma se acumulan, efectivamente, junto al ecuador galáctico.

formemente en las tres dimensiones del espacio, la distribución de intensidades seguirá una ley sencilla. El número de fuentes más intensas que un valor dado S es proporcional a $S^{-\alpha}$ donde α es 3/1. (Esta relación refleja que mientras el número de fuentes aumenta con el cubo de la distancia, su brillo aparente disminuye con el cuadrado de ésta.) La ley es válida incluso cuando no se conocen las direcciones, y tampoco viene afectada por posibles variaciones no sistemácas del brillo intrínseco con la distancia en el conjunto de los objetos implicados.

Sabemos, sin embargo, que nuestra galaxia es muy delgada en comparación con su diámetro. Por lo tanto, si las fuentes de las erupciones de rayos gamma siguen la misma distribución que las estrellas de alguna clase conocida, el número de fuentes no aumentará ya proporcionalmente al volumen cuando las distancias comiencen a exceder el espesor de la galaxia. El valor de α tenderá, de hecho, hacia la unidad. Una dificultad aneja a esta relación sencilla es que al ir disminuyendo el valor de S, la intensidad observada, algunos sucesos son ya demasiado débiles para ser detectados. Por debajo de un cierto valor de S se perderán todos los sucesos. Cuando se representa el valor de S en función del número de fuentes observadas, el valor decaerá por debajo de 3/2 y podrá llegar a ser cero, no a causa de la distribución de las fuentes, sino debido a nuestra incapacidad por detectar las débiles. Desgraciadamente, este efecto se manifiesta en seguida en los datos de Vela y no hemos podido llegar a ninguna conclusión definitiva a partir de ese estudio.

Un segundo método consiste en estudiar las pocas direcciones que tenemos y buscar claves de su distribución en el espacio, a partir de las cuales podamos calcular las distancias. Si representamos las ubicaciones de los siete sucesos para los que se tienen posiciones únicas, tenemos una distribución razonablemente aleatoria. Si las fuentes estuvieran bastante cerca del sol, digamos hasta 100 parsecs de éste, tal sería el tipo de distribución a esperar. Si estuviera de 10 a 20 veces más lejos, uno esperaría verlas empezar a acumularse a lo largo del ecuador galáctico (la Vía Láctea), siguiendo la distribución de las estrellas. Si estuvieran a distancias de 10.000 parsecs o más, detectaríamos la mayor parte de las fuentes de la galaxia; la asimetría, con la mayor parte de ellas situadas hacia el centro galáctico, sería obvia. No se observa tal asimetría.

Para distancias de más de 100.000 parsecs, las fuentes estarían fuera de la galaxia y debería esperarse de nuevo una distribución aleatoria, a menos que las fuentes estuvieran asociadas a galaxias cercanas, en cuyo caso estaríamos en condiciones de ver una correlación con dichas galaxias. Tenemos alguna información direccional sobre un total de 16 fuentes de erupciones de rayos gamma, incluidas las siete direcciones unívocas ya mencionadas. Ninguna de ellas corresponde, ni siquiera aproximadamente, a la posición de ninguna de las galaxias del Grupo Local, el cúmulo al cual pertenece nuestra galaxia. Podemos asegurar, pues, que si las fuentes son extragalácticas están a distancias de más de cuatro millones de parsecs.

Aunque sólo disponemos de siete direcciones unívocas, hay parejas de posiciones alternativas para nueve fuentes más. Debido a la orientación de los planos orbitales de los satélites observadores, los pares de puntos alternativos están, aproximadamente, a la misma latitud galáctica. Así pues, aunque no conocemos las longitudes galácticas, podemos hacer una estimación satisfactoria de las latitudes promediando, para ello, los valores alternativos. La distribución en latitud puede compararse entonces con la que se podría esperar a partir de una distrubución aleatoria. Esta comparación revela una apreciable concentración hacia el ecuador galáctico, pero el tamaño de la muestra es aún muy pequeño para sacar una conclusión firme.

Si suponemos que las fuentes de las erupciones de rayos gamma se acumulan junto al ecuador galáctico, debemos esperar que sus distancias tengan valores comprendidos entre unos centenares y unos miles de parsecs, dependiendo ello en gran medida de cuán grueso sea el disco determinado por las fuentes. No existe acuerdo entre los astrofísicos sobre si la prueba actual es convincente o no. Si uno cree que los puntos se acumulan en torno al ecuador, una distancia típica es 1000 parsecs. De lo contrario, las distancias son, o de 100 parsecs, o al menos cuatro millones de parsecs.

Modelos para las fuentes

En lo esencial todas las explicaciones de las erupciones de rayos gamma se fundan en el número limitado de hechos que acabamos de presentar. Algunos modelos se las arreglan incluso para no tener en cuenta o contradecir algunos hechos. En un artículo reciente, Melvin A. Ruderman, de la Universidad de Columbia, planteó adecuadamente el problema. "La mayoría de los astrofísicos teóricos" escribió, "sólo se desenvuelven bien en uno o dos modos normales. Debido a ello, a menudo nos retorcemos bastante fatigosamente para convencernos a nosotros mismos y a los demás de que las observaciones de un fenómeno nuevo caen dentro de nuestras especialidades de elección".

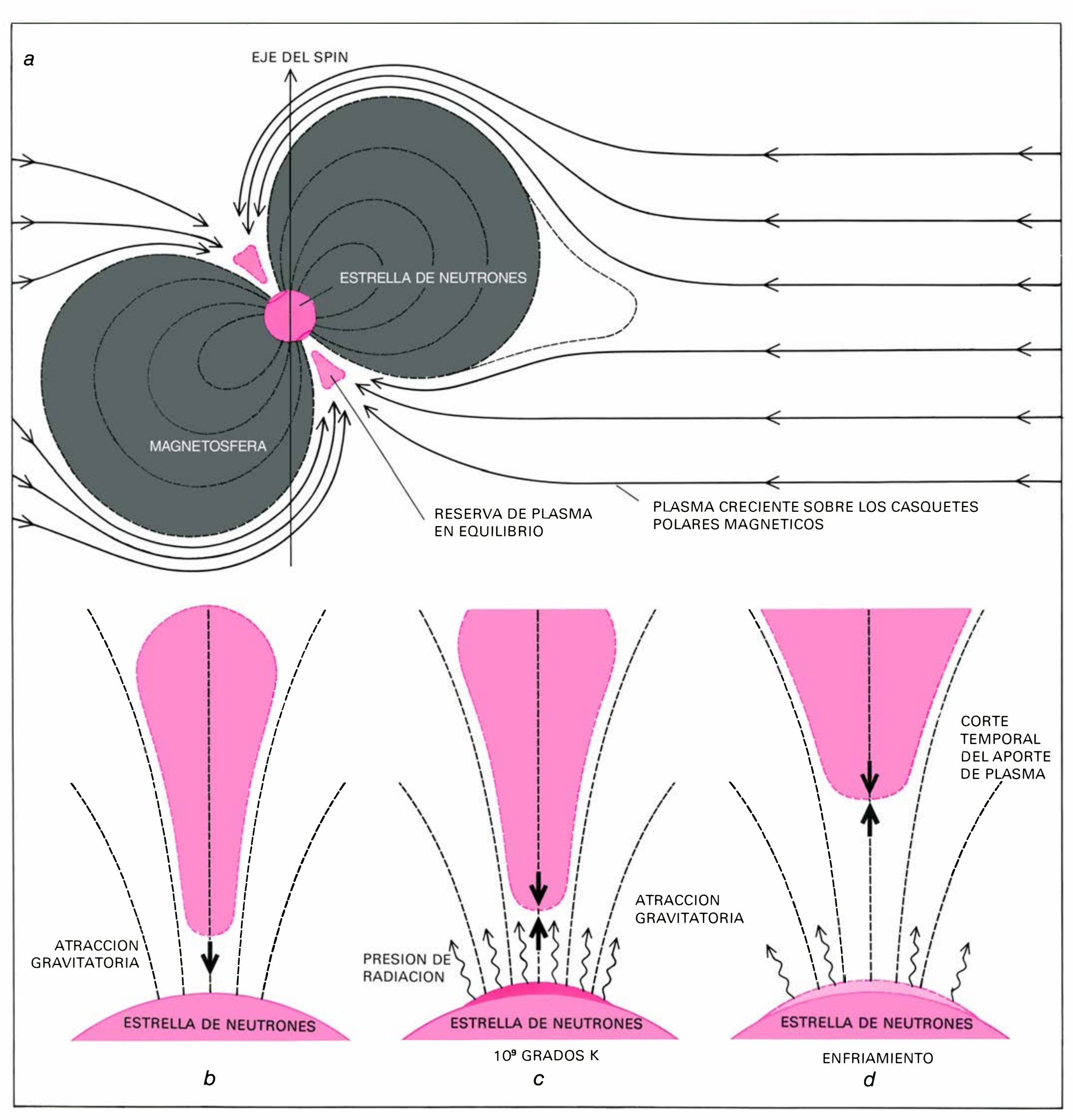
No sería posible describir con todo detalle los modelos de erupciones de rayos gamma, entre otras cosas porque muchos autores no han publicado los detalles. Describiremos sólo los modelos que llegan a predecir, al menos, algunas de las propiedades de las erupciones.

Hemos puesto de manifiesto que los modelos extragalácticos implican distancias de las fuentes de, por lo menos, cuatro millones de parsecs, o sea 13 millones de años-luz, e imponen, por tanto, un mínimo de 10^{47} ergios para la energía total radiada. Históricamente, la primera explicación propuesta fue un modelo, ideado por Stirling Colgate, del Instituto de Minería y Tecnología de Nuevo México, en el cual la onda de choque generada por una supernova se expansiona a velocidades relativistas, es decir, próximas a la de la luz. Cuando el núcleo de la estrella colapsa, iniciando así el proceso de supernova, se forma una onda de choque fuerte, esférica, que se propaga hacia el exterior a velocidades cada vez mayores, a medida que la densidad de la estrella va decreciendo. En la superficie de la estrella, la velocidad es relativista; los rayos X emitidos por el frente de choque, muy caliente, sufren corrimiento Doppler y forman un pulso de rayos gamma "blandos". Este mecanismo, el original de Colgate, dio pie a un intenso trabajo, observacional y teórico, en torno a los estadillos estelares transitorios, pero no parece que pueda explicar las erupciones de rayos gamma cósmicos. La energía total emitida es, probablemente, demadiado baja y la estructura fina de las erupciones queda sin explicar.

Un segundo modelo, propuesto por J. Cohen, de la Universidad de Pennsylvania, y por Reuven Ramaty, del Goddard Space Flight Center, recurre al colapso de una estrella enana blanca en estrella de neutrones. Es también un modelo extragaláctico, pues aunque un colapso tal podría ocurrir en cualquier galaxia, su frecuencia en la nuestra sería, con mucho, demasiado baja para dar cuenta del número de erupciones de rayos gamma conocidas. Verosímilmente, hay estrellas enanas blancas, dotadas de masa que capturan lentamente materia de sus proximidades. Cuando la fuerza gravitatoria sobre el núcleo de la estrella excede la capacidad de la materia para resistirla,

la estrella colapsa rápidamente, dando lugar a un objeto con la densidad típica de una estrella de neutrones (10^{14} gramos por centímetro cúbico, aproximadamente). La temperatura de la superficie de la estrella que colapsa puede alcanzar fácilmente los 10^9 grados Kelvin y emitir un relámpago de radiación gamma "blanda". Sólo se puede especular sobre la sucesión de acontecimientos que seguiría. Tal vez consistiera en una serie de colapsos de una capa, seguidos de algún tipo de rebote; pero, de nuevo, es difícil ver cómo podría dar cuenta el modelo de los picos más o menos aleatorios que aparecen en los registros conseguidos de las erupciones de rayos gamma.

Si la presente prueba en favor de la localización de las fuentes de erupciones de rayos gamma se corrobora, ambos modelos extragalácticos deberán desecharse. Además, Ruderman ha señalado que, salvo ciertas posibles excepciones,



EL MODELO DE "GLITCH" de estrella de neutrones recurre, para explicar las erupciones de rayos gamma, a cambios bruscos en la intensidad del campo magnético o en la rotación ("glitches") de estrellas de neutrones. Se cree que tales estrellas son residuos de supernovas: objetos de rotación rápida, de masas parecidas a las del sol, con un diámetro de menos de 50 kilómetros aproximadamente, y que poseen campos magnéticos muy intensos. Se puede suponer que algo del plasma de las proximidades de esas estrellas se recoja en los casquetes polares magnéticos, formando así una reserva de partículas cargadas mantenida en equilibrio por combinación de la atracción gravitatoria, la presión de radiación y la fuerza centrífuga *(a)*. Un "glitch" perturbaría el equilibrio, permitiendo así que un glóbulo de plasma se precipitara sobre los casquetes polares *(b)*, liberando una energía de hasta 100 MeV por partícula y creando radiación con una temperatura equivalente a mil millones de grados *(c)*. La posición de radiación rechaza al plasma y corta la radiación *(d)*. Puede repetirse hasta agotar la reserva de plasma.

resulta muy difícil explicar una luminosidad mayor que la de un cuerpo negro radiante extremadamente caliente, con una temperatura relacionada con la energía media de los fotones de los rayos gamma, la cual sería de unos mil millones de grados K. Sólo un objeto de mayor superficie podría tener una luminosidad más grande. Si el diámetro de las fuentes de rayos gamma es menor que 300 kilómetros, se sigue de ello que hay límite superior para la luminosidad absoluta; es de unos 2×10^{47} ergios por segundo. Lo cual sitúa la distancia máxima de las fuentes en unos cuatro millones de parsecs. Como ninguna fuente conocida tiene una posición que corresponda a la de ninguna galaxia dentro de los cuatro millones de parsecs, todos los modelos extragalácticos empiezan a parecer improbables (a menos, claro está, que la restricción sobre la luminosidad pudiera levantarse). Los modelos, sin embargo, no pierden interés, pues pueden predecir fenómenos de radiación que todavía no se hayan descubierto.

Modelos galácticos

Pasamos ahora a la amplia clase de modelos galácticos para las fuentes de las erupciones de rayos gamma. Las similitudes superficiales entre las erupciones de rayos gamma y las fulguraciones solares, que pueden emitir radiaciones X, han inspirado un cierto número de modelos de fulguración. En uno de ellos, propuesto por Floyd W. Stecker y Kenneth J. Frost, del Goddard Space Flight Center, las supuestas fulguraciones tienen lugar en estrellas enanas blancas con un fuerte campo magnético. En un modelo sugerido por Kenneth Brecher y Philip Morrison, del Instituto de Tecnología de Massachusetts, las erupciones de rayos gamma se generarían cuando los fotones de baja energía fueran "inflados" por electrones extremadamente energéticos, en fulguraciones producidas en estrellas de tipo más corriente. Otras hipotéticas fuentes de erupciones de rayos gamma hablan de colisiones entre materia y antimateria y entre partículas de polvo relativistas y fotones.

Cuando pasamos a las estrellas de neutrones, que pueden tener un campo magnético de un billón de gauss, resulta muy fácil encontrar varias fuentes posibles para la energía de una erupción de rayos gamma. Se cree que los púlsares giratorios, que emiten pulsos de radio a intervalos exactamente espaciados, desde una posición de segundo a varios segundos, son estrellas de neutrones. La energía del campo magnético de una estrella de neutrones sería mucho mayor que la de la erupción, para fuentes situadas a distancias moderadas, aunque el modo de apropiársela no está claro en absoluto. Se sabe que hay "glitches", o cambios bruscos en la velocidad de rotación de una estrella y neutrones, y tal vez podrían dar cuenta de la energía liberada. Franco Pacini, del Observatorio Austral Europeo, y Ruderman, han mostrado que las energías de las erupciones de rayos gamma coinciden en orden de magnitud con la energía perdida por una estrella de neutrones durante un "glitch", pero no es fácil explicar exactamente cómo se transformaría la energía de rotación en radiación gamma. En otros modelos construidos también a base de estrellas de neutrones, la fuente de los rayos gamma se adscribe a "terremotos estelares", a la actividad volcánica y a otros cambios bruscos de tamaño o forma. Por desgracia para los admiradores de esta clase de modelos, ninguna de las posiciones conocidas de erupciones de rayos gamma coincide con la de ningún púlsar conocido, aunque ahora se cree que hay un gran número de estrellas de neutrones además de las doscientas identificadas hasta la fecha a través de su emisión de pulsos de radio.

La clase de modelos que más atrae probablemente a los astrofísicos, entre los cuales nos contamos, es aquella en que la materia que procede de alguna parte, generalmente de una estrella normal, es atraída hacia la superficie de una estrella de neutrones o de una enana blanca, o hacia un agujero negro. Los agujeros negros son puntos, todavía hipotéticos, donde una estrella agotada ha colapsado hasta una densidad tan enorme que no puede escapar ninguna radiación, pero donde la materia que cae puede emitir fotones de mucha energía antes de perderse de vista. Un átomo de hidrógeno que cayera sobre una enana blanca puede liberar un millón de electrón-voltios (MeV), y otro que cayera sobre una estrella de neutrones puede liberar hasta 100 MeV; una cantidad de energía parecida puede liberarse junto a un agujero negro. La amplia gama de comportamientos que presentan las estrellas de rayos X ha sido explicada con éxito por medio de modelos en los que la estrella captura materia. Las fuentes binarias de rayos X se han llegado a entender, en base a ello, con bastante detalle.

Hemos propuesto una manera posible de obtener erupciones de rayos gamma a partir de la captura intermitente de materia por parte de una estrella de neutrones. En resumen, la idea es que hay glóbulos de plasma (un gas de electrones e iones) que se mantienen en equilibrio sobre los polos magnéticos de la estrella de neutrones. Cuando sucede un "glitch", los glóbulos aceleran hacia abajo, liberando rayos gamma durante el proceso.

Cygnus X-1: fuente posible

Durante un par de años hemos estado tentados por la posibilidad de que algunas erupciones de rayos gamma se originaran en Cygnus X-1, una poderosa fuente de rayos X y un sistema doble que se cree sea el más firme candidato a contener un agujero negro. Una erupción de rayos gamma observada el 12 de abril de 1972, tiene dos localizaciones posibles en el cielo, cada una de ellas dentro de un área de unos 20 grados cuadrados. Cygnus X-1 cae en medio de una de ellas. Nos apresuramos a señalar que un área de 20 grados cuadrados es tan amplia que la inclusión de Cygnus X-1 podría ser una pura coincidencia.

Otra erupción, con un flujo de energía y una duración parecidos, registrada el 15 de marzo de 1971, nos dio muchos dolores de cabeza. Ciertos datos suministrados por dos satélites Vela, combinados con algunos del OGO-5, nos confirman una amplia área posible que también contiene a Cygnus X-1, pero los cuatro tiempos de los satélites Vela no son compatibles.

Hay que ajustar dos de ellos en cerca de un segundo para obtener alguna dirección a partir de los datos. Si tuviéramos un error transitorio de cronometraje de exactamente un segundo en dos de ellos, obtendríamos dos buenas localizaciones, una de las cuales contiene de nuevo Cygnus X-1. El problema aún no ha sido resuelto.

Una mejor razón para asociar Cygnus X-1 con esas erupciones es que la radiación X habitual cambió drásticamente de carácter a mediados de marzo de 1971: desde entonces hemos encontrado dos erupciones más durante las raras ocasiones en que la radiación X de Cygnus X-1 ha cambiado. Podríamos decir que la prueba es suficiente para inculpar pero no para demostrar.

Las consecuencias de la asociación de erupciones de rayos gamma con Cygnus X-1, de probarse ésta, son fascinantes. No es sólo que Cygnus X-1 incluya posiblemente un agujero negro, sino que su distancia (2600 parsecs) se conoce con cierta precisión y que el objeto ha sido estudiado con gran detalle durante años. Conocida la distancia, se deduce que las erupciones tendrían un débito de energía

de unos 10^{40} ergios por segundo, o sea dos millones de veces el del sol en todas las longitudes de onda. La emisión de rayos gamma sería también unas 100 veces la energía del débito usual en rayos X de Cygnus X-1. El flujo de rayos gamma excede también en mucho el límite de Eddington del objeto: la cantidad de radiación estacionaria que expelería a la superficie de una estrella, en caso de mantenerse.

Es tentador sugerir mecanismos plausibles para las erupciones de rayos gamma, basados en la posible conexión con Cygnus X-1. Nosotros hemos sido de los que no han podido resistir la tentación. Un modelo, que nos parece viable, recurre a las interacciones entre dos corrientes de caída de plasma junto a un agujero negro. La interacción genera rayos gamma *(véase la ilustración de la página 53)*.

Debemos hacer notar algunas de las restricciones de tales modelos. Es necesario que expliquen de qué manera están relacionadas las erupciones con los dos estados (como mínimo) de emisión de rayos X del Cygnus X-1, la duración característica y la luminosidad de las erupciones y, en particular, por qué tienen lugar durante la transición entre estados. Si se demuestra que Cygnus X-1 contiene efectivamente un agujero negro, la posible asociación con las erupciones de rayos gamma suscita la intrigante posibilidad de que tales erupciones sean una característica de los agujeros negros, lo cual podría incluso ser camino útil para su detección.

Muchos modelos aquí mencionados quedan mejor caracterizados por el término "guión", que se ha puesto de moda recientemente entre los astrofísicos. Ello quiere decir que la información de detalle es tan escasa que las propuestas no merecen el calificativo de "modelos". No obstante, un buen guión puede, a veces, conducir a un buen modelo.

Estamos esperando ahora los primeros resultados de la combinación de los datos de tres nuevos satélites: el Helios-2, que lleva los primeros detectores especialmente diseñados para observar erupciones de rayos gamma, y un par de satélites del Laboratorio de Investigación Naval, llamados SOLRAD, para los que Los Alamos ha suministrado dos detectores de rayos gamma. Con suerte, podremos localizar erupciones de rayos gamma con precisión suficiente para poder identificarlos con objetos celestes visibles. Cuando esto suceda, vendrán rápidos progresos como los registrados en la comprensión de las estrellas de rayos X.

Hologramas de luz blanca

Existen varios modos de ver hologramas con luz incoherente normal, que hacen superflua cualquier dependencia de la luz coherente del láser y de la luz casi coherente de la lámpara de arco de mercurio

Emmett N. Leith

La técnica de la holografía, o fotografía por reconstrucción de frentes de onda, va a cumplir ahora los 30 años de existencia, tras su invención en 1947 por Dennis Gabor como un medio posible de mejorar el poder de resolución del microscopio electrónico. Sin embargo, hasta los primeros años de la década de los sesenta, no empezó la técnica holográfica a constituirse en centro de una amplia atención. Por aquella época Juris Upatnieks y el autor, trabajando en la Universidad de Michigan, introdujeron una serie de innovaciones que hicieron posible extender el principio de Gabor más allá de sus aplicaciones iniciales. Por ejemplo, el uso del método de haz de referencia fuera del eje permitió obtener imágenes holográficas de calidad superior a las obtenidas hasta entonces. Además, aprovechando la luz del láser, intensa y altamente coherente, nos fue posible obtener, por primera vez, imágenes holográficas sorprendentemente realistas de varios tipos de objetos reflectantes tridimensionales. El desarrollo posterior de la holografía ha ido profundamente ligado al del láser, por cuanto la demanda de mayores y mejores hologramas de escenas mayores y más ricas ha tendido a incrementar la necesidad de elevadas coherencias en las fuentes de luz empleadas para crear estas imágenes.

Al mismo tiempo, un deseo paralelo de hacer la holografía más práctica, y por tanto más universal, ha inducido a la búsqueda de métodos que reduzcan las exigencias de coherencia del proceso. En la pasada década se hicieron notables progresos por lo que respecta a ambos objetivos. En efecto, en exposiciones públicas recientes se han mostrado hologramas que pueden observarse con luz incoherente, o blanca, luz solar, es decir, a partir de una lámpara corriente de incandescencia, y no sólo con la ayuda de la luz coherente de un láser. Mucha gente ha quedado sorprendida al ver que estos hologramas que durante mucho tiempo, anduvieron asociados con la tan compleja y costosa tecnología del láser, podían observarse con estas fuentes de luz comunes.

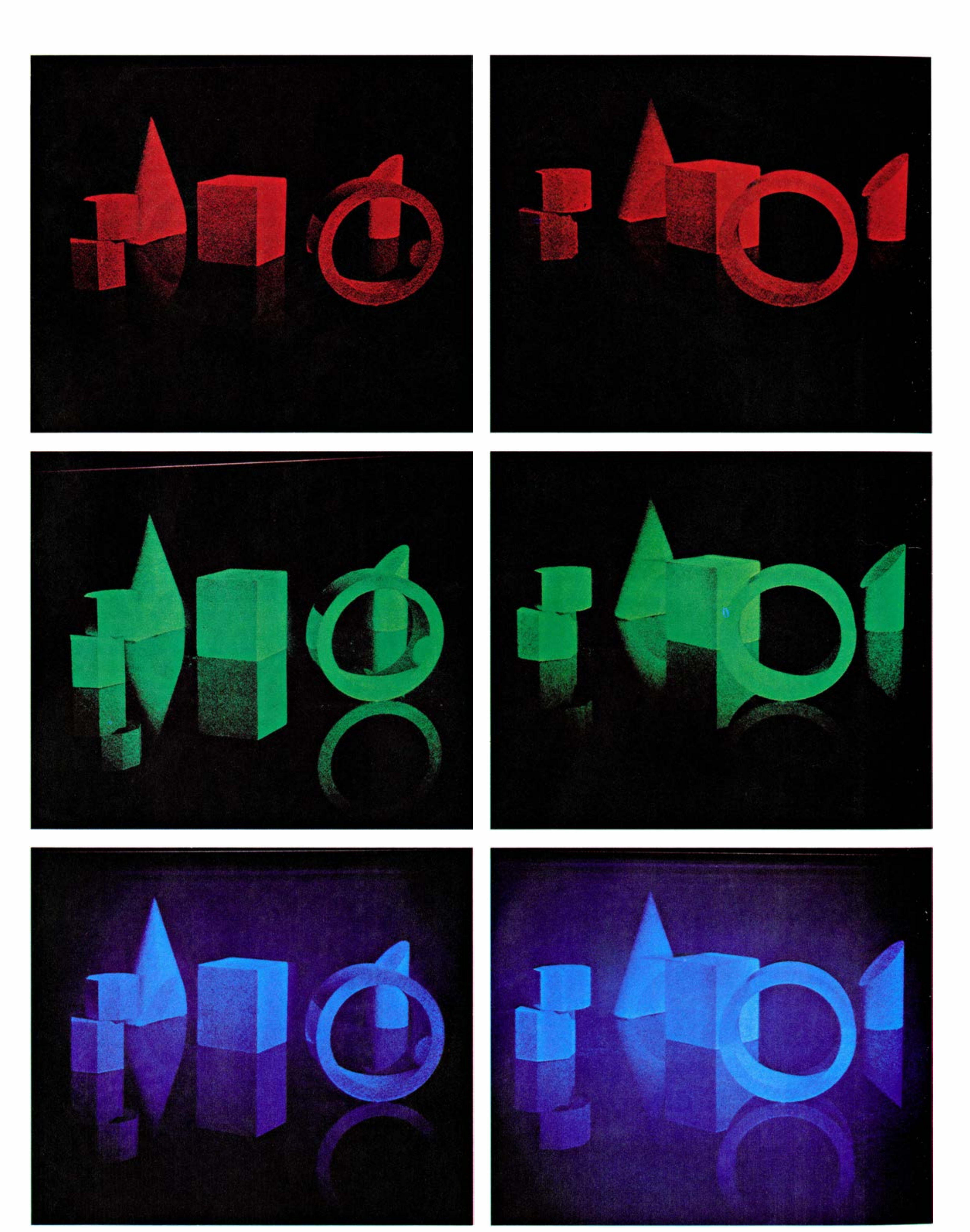
En este artículo explicaré por qué la luz coherente ha desempeñado un papel tan importante en la holografía; hablaré después de cómo difieren las exigencias de coherencia para las distintas clases de hologramas, y, finalmente, describiré algunos avances que han contribuido a la reducción sustancial de esta exigencia, tanto para ver como para hacer hologramas.

En su forma mejor conocida, la holografía es un proceso relativamente directo. La luz coherente que procede de un láser se divide en dos rayos. Uno se usa para iluminar un objeto, y la parte del rayo que es reflejada por el objeto incide en una placa fotográfica. El otro rayo se dirige directamente a la placa por medio de un espejo. La luz procedente del espejo, llamada rayo de referencia, se combina en la placa con la luz reflejada por el objeto para formar una imagen compleja de interferencia. La placa revelada que tomó aquella imagen de interferencia es el holograma. Cuando se ilumina un holograma de este tipo con sólo el rayo de referencia, los rayos de luz que atraviesan la placa son transmitidos o absorbidos selectivamente de manera que creen en el haz emergente una componente que duplica las ondas luminosas originales, que fueron reflejadas desde el objeto a la placa cuando se formó el holograma. Un observador que reciba estas ondas, las percibirá como si emanasen del objeto original; en efecto, él "verá" el objeto como si realmente estuviera allí *(véase la ilustración de la pág. 61)*.

La exigencia de coherencia surge porque el holograma es una grabación de la interferencia de dos ondas; en general (pero no siempre), la luz coherente produce unas imágenes de interferencia y la incoherente no. Esta exigencia ha dificultado seriamente el desarrollo de la holografía, en concreto con objeto de exhibirla, ya que las fuentes de luz coherente no sólo son mas caras sino también más incómodas. Asi pues, cualquier reducción en el requisito de coherencia, principalmente en el proceso de visualización, tendría una considerable importancia para la holografía de exhibición.

¿Cuál es la diferencia entre luz coherente y luz incoherente? Hay dos tipos de coherencia: temporal y espacial; se dice que la luz es coherente cuando posee ambos tipos. La luz coherente temporalmente es la luz monocromática: tiene solamente una longitud de onda. La luz coherente espacialmente es la luz que procede de una fuente puntual o es capaz de ser localizada en un punto. Hoy en día la fuente usual de luz coherente es el láser. Sin embargo, la luz coheren-

FOTOGRAFIAS DEL HOLOGRAMA EN "ARCO IRIS" construido por Stephen Benton, de la Polaroid Corporation, y Fritz Goro *(página opuesta)*. **Aunque el holograma aparece en tres colores diferentes, se fotografió con luz blanca transmitida por el condensador de un microscopio con lámpara de filamento incandescente de tungsteno. Los tres colores diferentes resultan en virtud del principio de que el color del holograma cambia a medida que la cámara (o el ojo) sube o baja; el holograma en cuestión se fotografió con la cámara en tres posiciones verticales distintas. Cuando la cámara se mueve de un lado a otro puede verse un rasgo característico de todos los hologramas: el paralaje. Este es el desplazamiento aparente de objetos en la escena tridimensional, cuando se contempla el holograma desde ángulos diferentes. Tal desplazamiento se advierte en cada uno de estos tres pares de fotografías. En la página 67 se da el diagrama de construcción.**







HOLOGRAMA DE LUZ BLANCA DE DENISYUK, realizado por el investigador ruso Yu. N. Denisyuk, fotografiado desde dos ángulos verticales distintos para demostrar no sólo el paralaje vertical conseguido con este método sino también el cambio abrupto en el color de la imagen percibida cuando se contempla desde dos direcciones diferentes. Los hologramas del tipo Denisyuk se ven con luz reflejada, obtenida a partir de una fuente puntual, en este caso una pequeña linterna de luz incandescente *(véase la figura inferior de la página 66).* **El objeto es un medallón en relieve.**

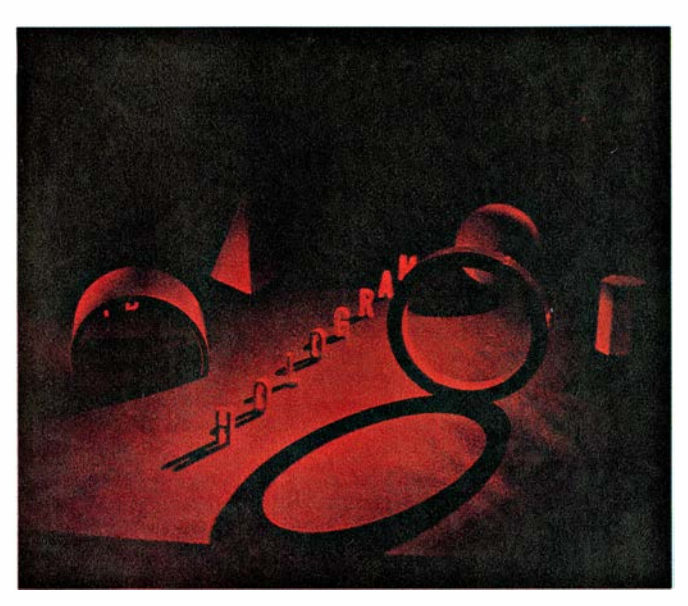
te puede obtenerse también a partir de fuentes incoherentes, como era habitual antes de que se inventase el láser. La coherencia espacial puede conseguirse, por ejemplo, situando una abertura muy pequeña delante de la fuente, de modo que la luz proceda toda de ese punto. La coherencia temporal, por el contrario, puede conseguirse situando un filtro de color delante de la fuente, de forma que sólo se transmita la luz de una banda espectral estrecha. Ambos procesos implican malgastar la mayor parte de la luz; de aquí que, para obtener una pequeña cantidad de luz coherente, sea necesaria una fuente de luz incoherente extremadamente intensa. Antes del láser, la lámpara de arco de vapor de mercurio constituía una de las mejores fuentes disponibles de luz coherente. El arco es intenso y generalmente tiene diámetros típicos de unos pocos milímetros. Además, la luz se emite en varias bandas espectrales estrechas y no en un espectro continuo. Así pues, las técnicas empleadas para obtener luz coherente a partir de una lámpara de arco de mercurio producen menos pérdidas de luz que si la fuente fuera, por ejemplo, un filamento incandescente de tungsteno.

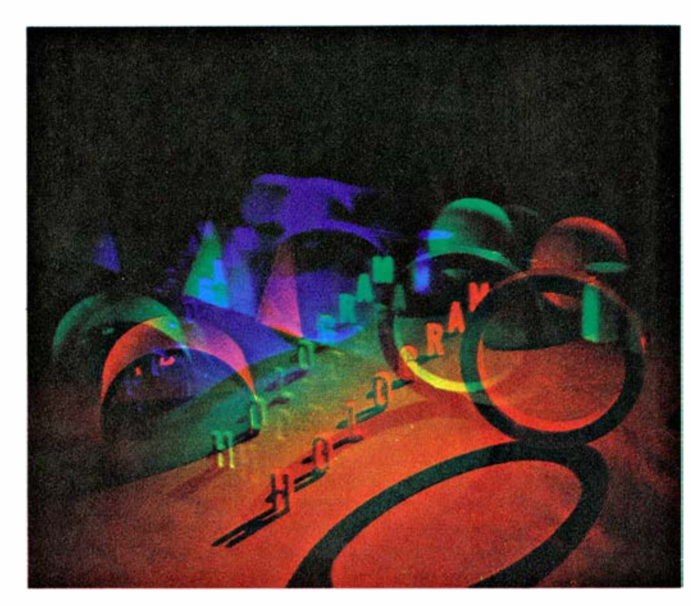
Los hologramas originales de Gabor se realizaron y visualizaron con una lámpara de arco de mercurio. Los objetos eran bastante simples: diapositivas con rótulos opacos y sin escala de grises; una parte de la luz incidente se hacía pasar a través de superficies abiertas, relativamente amplias, y servía así de haz de referencia. La luz, en las proximidades de las líneas oscuras que formaban las letras, era difractada o dispersada, y por lo tanto se esparcía de manera que las letras resultaban irreconocibles, vistas en el plano donde se grababa el holograma. Estos hologramas no parecen impresionantes para los estándares actuales, pero constituyeron una demostración convincente de un nuevo e interesante principio de la óptica. En 1971, Gabor recibió el Premio Nobel de física en reconocimiento a su invento.

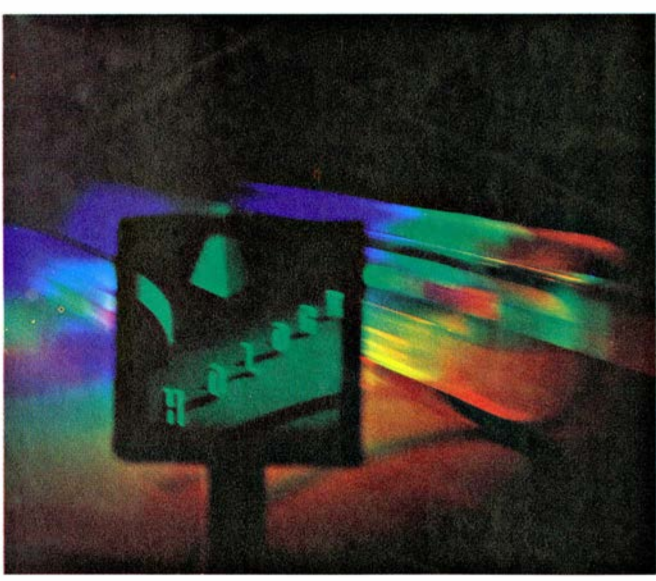
Las exigencias de coherencia de los primeros hologramas de Gabor no eran muy importantes. En general, estos requisitos de coherencia pueden calcularse a partir de consideraciones bastante elementales en lo que respecta a la realización y visualización de los hologramas. La luz dispersada por cada elemento resolutivo (el punto más pequeño distinguible en el objeto) interfiere con el haz de referencia no dispersado para dar lugar a una imagen de difracción de "ojo de buey", que consiste en una

serie de anillos concéntricos, brillantes y oscuros alternativamente; éstos se van haciendo más finos a medida que nos vamos alejando del centro. La luz que procede de cada elemento resolutivo se dispersa en la forma dicha; las imágenes de los puntos próximos se superponen, formando imágenes más complejas. El holograma es una superposición de muchas imágenes elementales, una por cada elemento resolutivo del objeto.

La escala de este diagrama de ojo de buey es proporcional a la longitud de onda de la luz. Si la luz es policromática, esto es, si hay más de una longitud de onda, se puede suponer que cada componente con una longitud de onda dada forma su propio conjunto de imágenes de difracción, una por cada elemento resolutivo del objeto (*véase la figura superior de la pág. 64*). Debido a que las imágenes correspondientes a diferentes longitudes de onda difieren en tamaño, un haz con una amplia gama de longitudes de onda emborronará el diagrama resultante, siendo las líneas más finas las más borrosas. La anchura aceptable de la banda espectral se hace más estrecha al aumentar la distancia entre el objeto y el holograma, o al reducir el tamaño del elemento de resolución. La relación entre el tamaño de la imagen de difracción y el tamaño del ele-

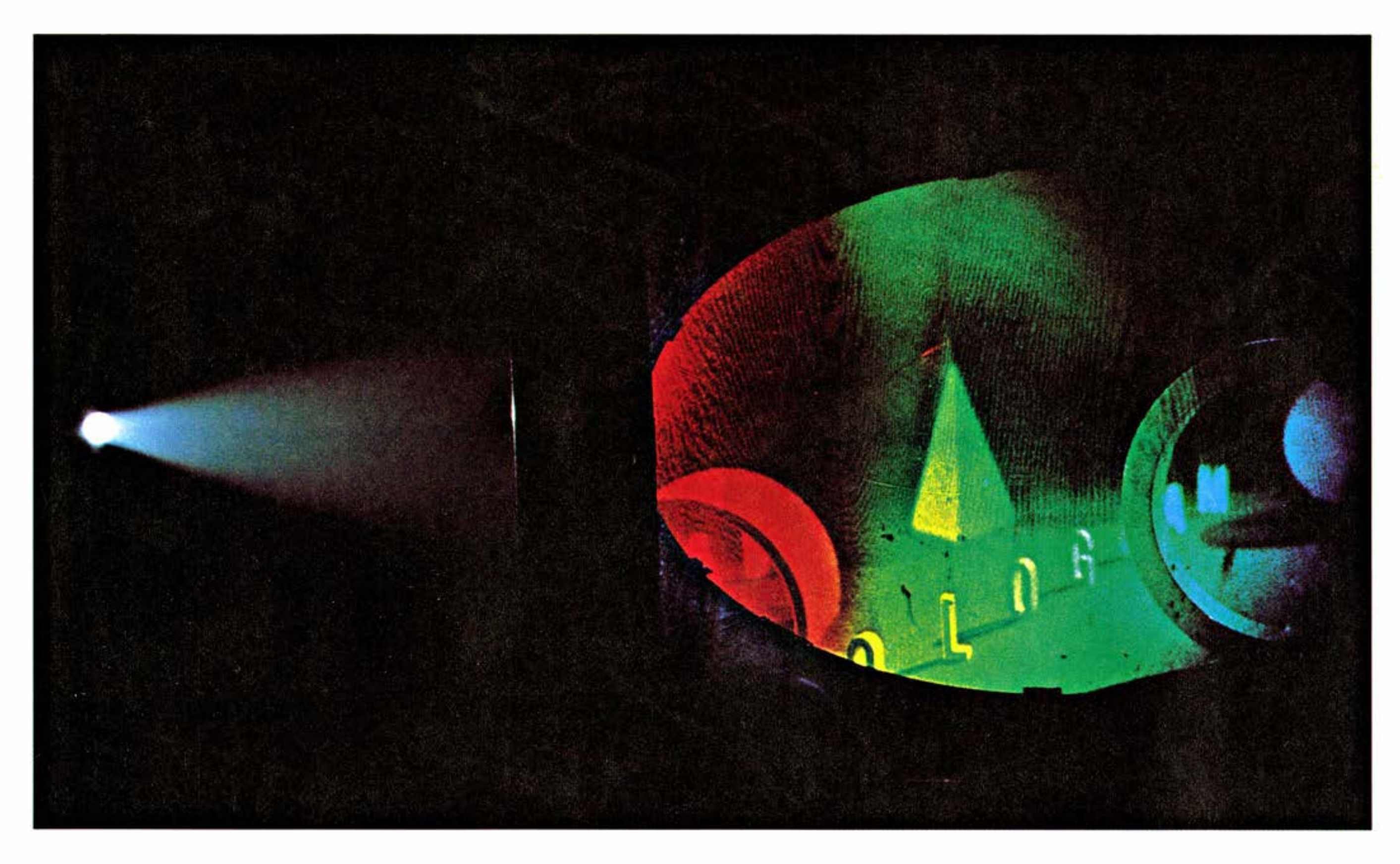








HOLOGRAMA CONVENCIONAL construido por Goro en el laboratorio del autor en la Universidad de Michigan. Se fotografió con luz transmitida desde 4 focos de luz diferentes para ilustrar la importancia de la exigencia de coherencia en el proceso holográfico tradicional. La fotografía superior, izquierda, muestra el holograma visto con la luz roja altamente coherente de un láser de helio-neón. La parte superior, derecha, muestra el holograma visto con luz policromática menos coherente emitida a varias longitudes de onda discretas por dos focos diferentes. La imagen roja está formada con luz del láser de helio-neón; las azules, verdes y ámbar con luz de una lámpara de arco de mercurio. El desplazamiento de las imágenes correspondiente a las diferentes longitudes de onda se debe a que este tipo de hologramas, llamados hologramas fuera del eje o de frecuencia portadora, actúan como una red de difracción, desviando la luz proporcionalmente a su longitud de onda y, por tanto, descomponiendo el haz de luz incidente en sus colores componentes. La fotografía de la parte inferior, izquierda, es una exposición doble; se hizo cubriendo primero una pequeña parte del holograma con luz blanca emitida por una lámpara de arco de circonio. La difuminación de los colores se produce porque cada longitud de onda en el espectro continuo de la luz blanca forma su propia imagen, y todas las imágenes están ligeramente desplazadas entre sí. La parte cubierta del holograma se reemplazó después por un filtro verde de banda estrecha, y se cubrió el resto del holograma. La segunda toma, hecha con ayuda del mismo foco de luz blanca del arco de circonio, revela nítidamente una imagen verde de la escena original. La fotografía de la parte inferior, derecha, vuelve a mostrar el holograma iluminado con el foco de luz blanca del arco de circonio, pero con una red especial intercalada entre el foco y el holograma a fin de compensar la dispersión cromática. La imagen aparece de esta forma con bastante nitidez, en el centro, donde la compensación es mejor.



HOLOGRAMA EN ARCO IRIS DE LA MISMA ESCENA que aparece en las fotografías del holograma convencional expuesto en la página anterior. Se consiguió asimismo en el laboratorio del propio autor. Cuando se contempla un holograma de este tipo con luz blanca transmitida desde una distancia de unos 90 cm, la imagen holográfica resultante varía de color a lo ancho del holograma (de ahí la expresión de holograma en arco iris). El foco de luz blanca fue aquí una lámpara de filamento de tungsteno. El humo hace visible el haz de luz.

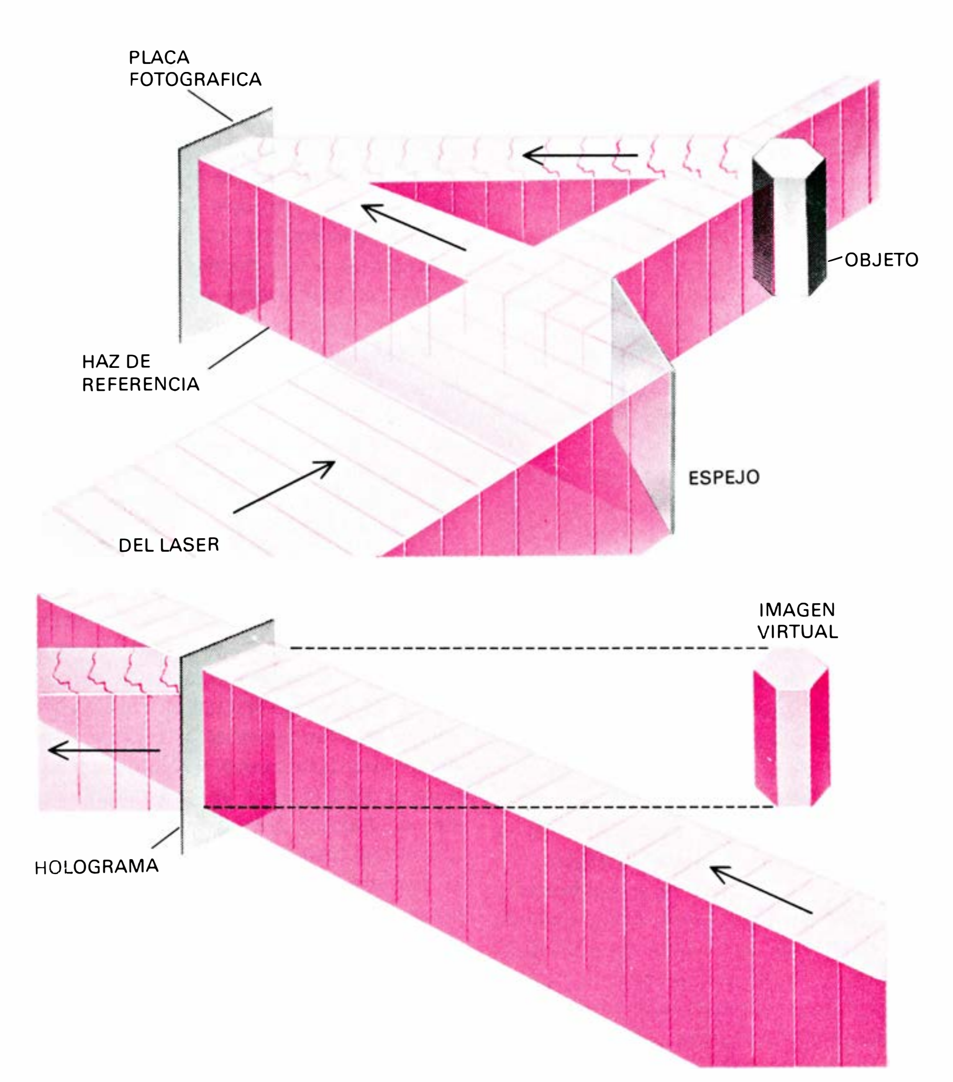
PRIMER PLANO DEL HOLOGRAMA EN ARCO IRIS que se muestra en la fotografía superior de esta misma página, obtenido con la misma fuente de luz blanca de filamento de tungsteno, pero en este caso desde una distancia de unos 30 centímetros; el color aparece uniforme en toda la imagen. Sin embargo, si se desplazara lateralmente la posición de mira, el color cambiaría pero seguiría siendo uniforme. La escena original que aparece en los hologramas representados en esta misma página y en la anterior tiene 50 centímetros de ancho.

mento resolutivo se llama a veces razón de expansión, ya que cada uno de estos elementos se difunde por tal relación para formar el holograma. Puede demostrarse que, para que semejante emborronamiento alcance valores mínimos, la anchura de la banda espectral debe ser menor que cuatro veces la longitud de onda media dividida por la razón de expansión.

Por ejemplo, si se toma como objeto una transparencia cuyo detalle más fino sea de 0,01 milímetros, se ilumina con un haz de luz que tenga una banda de longitudes de onda centrada en los 5000 angstroms y se elige una separación entre el objeto y el holograma tal que se obtenga una razón de expansión de 50, se encuentra que es aceptable una dispersión de 400 angstroms en la longitud de onda. Ello representa la octava parte del espectro visible completo, y un haz que cubriese tal anchura de banda difícilmente podría llamarse coherente. Una lámpara de arco de mercurio ordinaria emite una línea de color verde intenso de uno o dos angstroms de anchura, por lo que tiene más coherencia temporal de la necesaria para construir hologramas.

Por vía alternativa, se puede describir la exigencia de coherencia temporal en términos de longitud de coherencia, una cantidad básica que representa esencialmente la distancia a que la frecuencia de vibración de una onda se mantiene con precisión: las ondas monocromáticas tienen largas longitudes de coherencia, mientras que por el contrario las ondas policromáticas no. En consecuencia, si se tiene que dividir un rayo de luz en dos partes, y más tarde recombinarlas, los efectos de interferencia surgirán siempre que la diferencia entre los dos caminos ópticos no exceda la longitud de coherencia de esa luz. Aunque este modo de describir la coherencia temporal parece no tener relación con el anterior, basado en la superposición de las franjas de interferencia formadas por diferentes longitudes de onda, de hecho, las dos formulaciones son equivalentes.

La etapa de reconstrucción, o visión, de la holografía es esencialmente una reconstitución de la etapa de formación del holograma, y se aplican las mismas consideraciones de coherencia temporal. Cuando un haz de rayos de luz colimados, o paralelos, incide en el holograma, cada imagen de difracción elemental se apropia de una porción de la luz incidente, convirtiéndola en un frente de onda esférico que parece emanar del elemento objeto original; la imagen formada se llama imagen virtual. La imagen de difracción elemental grabada tiene un comportamiento muy parecido al de una lente que focaliza la luz incidente en un punto de la imagen; ciertamente, la estructura de ojo de buey del diagrama de difracción es similar al que produciría una lente en la zona de Fresnel, dispositivo que se emplea frecuentemente en los cursos de óptica. El tamaño del elemento de resolución de la imagen virtual es exactamente el mismo que el tamaño del elemento resolutivo del objeto original, a menos que algunas imperfecciones en el proceso hayan degradado la imagen. La razón del tamaño del diagrama de difracción al tamaño del elemento resolutivo de la imagen se llama razón de compresión: es, por supuesto, exactamente igual a la razón de expansión.



LA FORMA DE HOLOGRAFIA MEJOR CONOCIDA, representada aquí, se basa en la luz coherente del láser. En la fase de construcción *(arriba)*, **se divide un rayo láser en dos partes. Una se usa para iluminar el objeto, y parte de la luz reflejada incide sobre una placa fotográfica. La otra parte se dirige a la placa ayudándose de un espejo. La placa registra las complejas imágenes producidas por la interferencia de los dos rayos. En la fase de reconstitución** *(abajo)*, **se ilumina el holograma sólo con el rayo de referencia, originando en el rayo emergente un duplicado exacto de los frentes de onda reflejados desde el objeto. Se puede producir el rayo de referencia con un láser o con un foco casi coherente, tal como una lámpara de arco de mercurio. La imagen percibida mirando por el holograma se conoce como imagen virtual.**

La necesidad de monocromaticidad en el proceso de visualización surge debido a que la curvatura del frente de onda, y por lo tanto la distancia al punto focal, es proporcional a la longitud de onda de la luz. Por eso, si se usa luz de muchas longitudes de onda, cada longitud de onda forma una imagen en una posición diferente. En consecuencia, en cualquier plano que uno enfoque el sistema de visión (el ojo o la cámara fotográfica) se puede ver una imagen en el foco y otras muchas imágenes, formadas con otras longitudes de onda, algo separadas del foco. Este defecto en la imagen resultante se llama dispersión longitudinal. La anchura de banda espectral de la fuente luminosa deberá ser lo bastante estrecha como para que tales efectos de emborronamiento sean despreciables. De he-

cho, se puede demostrar que la anchura de la banda espectral no deberá ser mayor que la longitud de onda media multiplicada por la razón de compresión, que es exactamente el criterio seguido en la construcción del holograma.

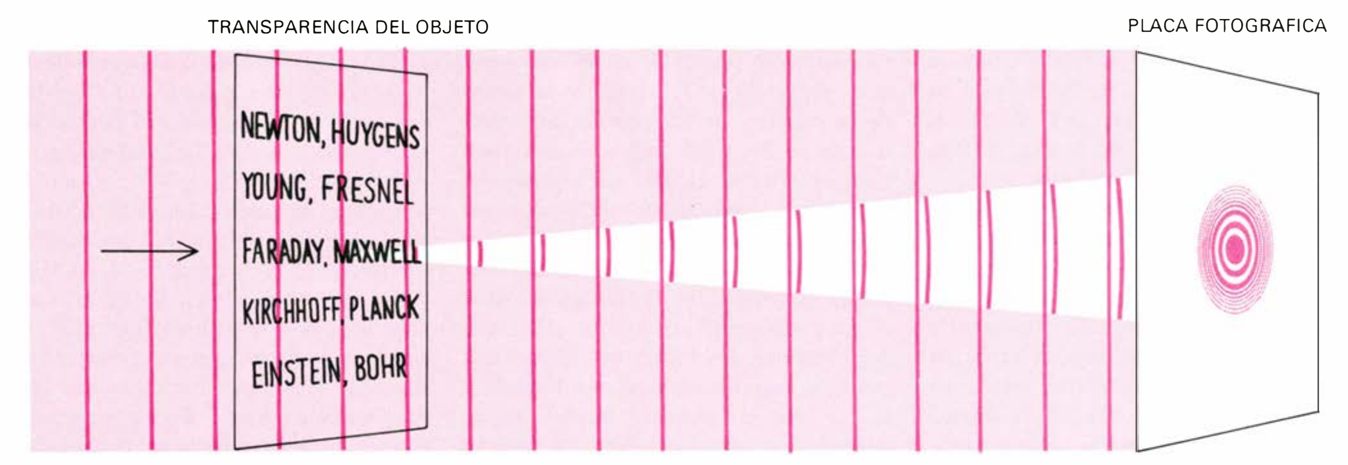
La razón de compresión (o la razón de expansión) es el único factor que entra en la determinación de los requisitos de coherencia temporal. El mismo tamaño del objeto resulta irrelevante; por tanto, la luz de coherencia limitada no limita en absoluto el tamaño del objeto que puede ser holografiado.

La coherencia espacial, el otro aspecto de la coherencia, está relacionada con el tamaño de la fuente luminosa. Una fuente extensa puede considerarse como formada por muchas fuentes puntuales separadas ligeramente, cada una de las cuales forma una imagen de difracción por cada elemento resolutivo del objeto *(véase la ilustración inferior de la pág. 64)*. Todos estos diagramas son proyectados a lo largo de la dirección de propagación del haz de luz incidente. Así pues, los diferentes elementos resolutivos de la fuente luminosa formarán imágenes de difracción de ojo de buey ligeramente desplazadas: si el tamaño de la fuente de luz es grande, la imagen de difracción resultante aparecerá borrosa por culpa de la fuente completa, y lo será particularmente en los bordes, en donde el espaciado de las franjas es más fino. Obviamente, la fuente de luz empleada para realizar un holograma deberá ser lo suficientemente pequeña como para que esta clase de difusión resulte inapreciable. De modo semejante, en el proceso de visualización, una fuente luminosa extensa lleva a una difuminación de la imagen; por tanto, la exigencia de coherencia espacial es la misma que en la formación del holograma.

Nuestro grupo de Michigan se dedicó a la holografía óptica en 1960, cuando intentamos mejorar la calidad de la imagen, que era por aquel entonces bastante pobre. Un problema básico con que había que enfrentarse era que cada elemento resolutivo del holograma forma dos ondas de curvatura igual y opuesta; la onda adicional es convergente, formando una imagen delante del holograma llamada imagen real. Se puede observar o la imagen virtual o la imagen real, pero siempre con la otra imagen desenfocada en el fondo. Este defecto se debe a que en el proceso de grabado se pierde el sentido de la curvatura del frente de onda. (Una onda esféricamente divergente y otra esféricamente convergente formarían la misma imagen de difracción.) El proceso de reconstrucción resuelve esta ambigüedad al producir dos ondas, una convergente y otra divergente, pero la calidad de la imagen resulta consiguientemente degradada. También se presentan otros defectos, que conducen a componentes adicionales de luz extraña.

Upatnieks y el autor solucionaron estos problemas mediante un procedimiento muy empleado en tecnología de la comunicación. Recurriendo a una técnica semejante en cierto modo a la modulación de una onda de radio sobre una frecuencia portadora, se provocó que una porción del haz luminoso incidiese oblicuamente en la placa. En consecuencia, las franjas de interferencia se hicieron más finas y numerosas, y, el holograma, mucho más parecido a una retícula de difracción ordinaria de líneas paralelas. Del mismo modo que una red de difracción dispersa un haz de luz incidente en muchos rayos separados (llamados órdenes) que viajan en diferentes direcciones, nuestros hologramas produjeron la difracción en varias direcciones. Los rayos que forman las dos imágenes se propagan ahora en dos direcciones diferentes y por tanto se separan. Todavía más, la otra serie de rayos extraños se propaga en direcciones diferentes. Resultando así que cada una de las dos imágenes, real y virtual, estaba libre de efectos extraños y tenía una calidad muy mejorada.

Con muchas más franjas, y un espaciado más fino entre éstas, se podría suponer que se requería una mayor coherencia de la fuente de luz usada para construir tales hologramas. Por otra parte, si el holograma actúa de forma semejante a una red de difracción, ¿no podría el holograma tender a dispersar la luz incidente en sus colores componentes como hacen las redes de difracción? (Las redes de difracción desvían la luz transmitida proporcionalmente a su longitud de onda; así pues, dentro de cada orden, la componente roja, por ejemplo, es desviada mucho más que la azul.) Por lo tanto parecía razonable esperar que este tipo de hologramas, generalmente llamado holograma de frecuencia portadora u holograma fuera del eje, formaría para cada longitud de onda una imagen desplazada ligeramente en sentido lateral de la imagen formada por una longitud de onda ligeramente diferente. *(véase la ilustración superior de la pág. 65)*. Para una misma banda bastante estrecha de longitudes de onda, la imagen resultaría muy difuminada. Este defecto en la imagen, llamado dispersión



METODO HOLOGRAFICO ORIGINAL, inventado en 1947 por Dennis Gabor, que empleó transparencias simples con rótulos opacos como objetos y como foco luminoso una lámpara de arco de mercurio. El diagrama de la izquierda, representado en esta misma página, muestra cómo se hizo el holograma. La mayor parte de la luz pasó a través de la transparencia y sirvió como haz de referencia. Sin embargo, en torno a las líneas oscuras que formaban las letras, una parte de la luz transmitida se difractó o dispersó en forma de frente de onda esférico por todos los elementos de resolución (o punto resolutivo más pequeño) sobre la transparencia. La interferencia de los frentes de onda difundidos desde cada elemento de resolución con los frentes de onda no difundidos del haz de referencia crearon en la placa fotográfica una imagen de

lateral, existe con independencia de la dispersión longitudinal descrita anteriormente. Podría parecer que las exigencias de coherencia que trae consigo nuestra holografía perfeccionada serán mayores que las requeridas por el tipo de Gabor, e incluso que quizás sólo podrían conseguirse hologramas con luz procedente de un láser.

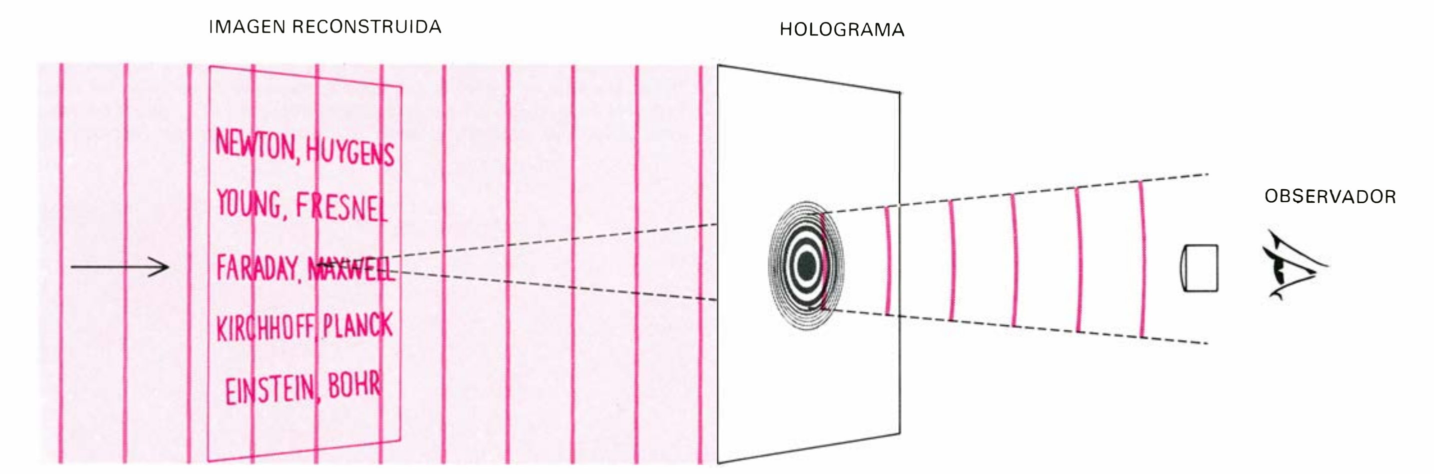
Ambas suposiciones son incorrectas. Nuestros primeros hologramas del tipo frecuencia portadora, fueron hechos realmente en la época anterior al láser y sin más medios que una fuente de luz convencional de arco de mercurio. Puesto que las exigencias de coherencia para la holografía básica no son de suyo muy grandes, resulta incluso que un aumento de varias veces tales exigencias queda dentro de la capacidad de la lámpara de arco de mercurio. Además, este incremento no es fundamental y puede evitarse fácilmente.

Supongamos, por ejemplo, que en el proceso de reconstrucción se coloca, entre la fuente y el holograma, una red de difracción con un espaciado igual a la separación media entre las franjas del holograma, y supongamos asimismo que para iluminar el holograma se utiliza precisamente el orden que desvía la luz en sentido opuesto al del holograma. Las dispersiones entonces se compensarían. Desde el punto de vista del holograma. cada longitud de onda componente de la luz incidente incidiría desde una dirección ligeramente diferente, siendo el desplazamiento angular exactamente el necesario para hacer que las imágenes formadas por las diferentes longitudes de onda coincidan. Si se hacen los ajustes adecuadamente, las exigencias de coherencia pueden reducirse a sólo las requeridas por el método de Gabor *(véase la ilustración central de la pág. 65)*.

En la construcción del holograma se aplica el mismpo principio. Para que los haces de luz policromática interfieran, el espaciado de las franjas de interferencia tiene que ser proporcional a la longitud de onda. Por tanto, si las franjas están alineadas en el lugar donde los rayos se superponen, se irán desalineando progresivamente a medida que se alejen del lugar; como consecuencia, se produce un número desajustado de franjas de interferencia. Sin embargo, situando una red entre la fuente y el objeto se puede predispersar la luz para hacer que el espaciado de las franjas sea el mismo para todas las longitudes de onda; el número de franjas que entonces se puede obtener es independiente de la longitud de onda. La red compensaría la dispersión lateral, reduciendo así las exigencias de coherencia, tanto para la construcción como para la observación del holograma, a las requeridas según el esquema original de Gabor para la holografía. El sistema descrito en nuestro ensayo original sobre la holografía incorporaba una red como la anterior, y en efecto se compensó las longitudes de onda, aunque no hubiera sido este nuestro objetivo, y la simplicidad del objeto hiciera tal compensación innecesaria. Robert E. Brooks, L. O. Heflinger y Ralph F. Wuerker, de TRW Inc., sugirieron y demostraron métodos similares de hacer los hologramas con láser de pulsos, pero entonces los láser de pulsos carecían de coherencia para realizar este experimento.

Hay algunas otras técnicas que reducen las exigencias de coherencia de nuevas y más avanzadas formas de holografía a las del método de Gabor, circunstancia esta que sugiere que las exigencias originales de coherencia constituyen un mínimo básico irreducible. Pero esta suposición resulta también falsa; de hecho, se pueden reducir más todavía las exigencias de coherencia. Por ejemplo, al iluminar un holograma de Gabor se podría interponer una placa de zonas de Fresnel entre el holograma y la fuente de iluminación. Se podría entonces tomar la imagen real formada por la placa de zonas como fuente para iluminar el holograma. (La placa de zonas forma una imagen dispersada longitudinalmente de la fuente, con el violeta que posee la longitud de onda más corta situado en la posición más alejada de la placa.) Con esta disposición se compensa la dispersión longitudinal del holograma, haciendo que la imagen de cada longitud de onda se forme a la misma distancia del holograma *(véase la ilustración inferior de la pág. 65)*. De aquí se deduce que un holograma de este tipo podría ser visto con luz blanca. Un argumento análogo podría establecerse en cuanto al proceso de formación del holograma, lo que nos lleva a su vez a la sugerencia de que también sería posible realizar hologramas con luz completamente blanca. Ciertamente, si se usaran combinadas una placa de zonas y una red, podría compensar simultáneamente tanto la dispersión longitudinal como la lateral, haciendo por tanto posibles que los hologramas con el haz fuera del eje, de alta calidad, se construyeran y visualizaran con luz totalmente blanca.

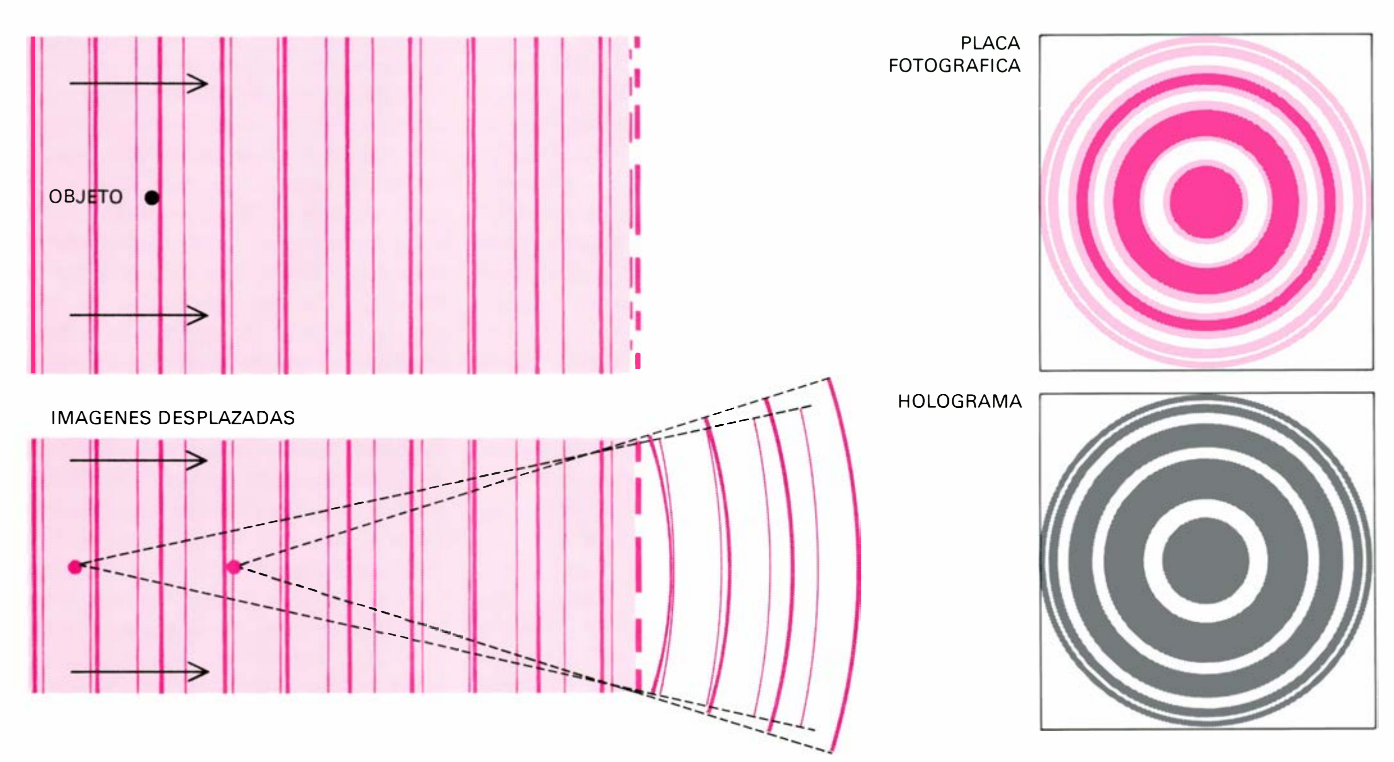
Existe un aspecto histórico interesante de esta hipótesis. En 1955, trabajando en Michigan con mis colegas en sistemas de formación de imágenes de radar, se nos



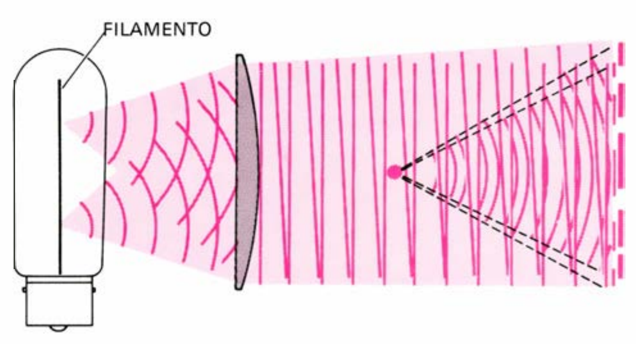
ojo de buey muy semejante a una placa de difracción de zonas de Fresnel. El holograma resultante es una estructura de recubrimiento de muchas imágenes, una por cada elemento de resolución del objeto original. El diagrama de la derecha, que se representa en la página siguiente, muestra como se veían los hologramas revelados. La difracción de una parte del rayo de referencia transmitido por una imagen de interferencia en ojo de buey se observa al recrear una imagen virtual de un elemento de resolución del objeto original. Los nombres de los científicos que aparecen en el holograma representado en estas ilustraciones estaban representados en el holograma original de Dennis Gabor, en *Proceedings of the Royal Society* en 1949. Este se proponía en su investigación mejorar el poder de resolución del microscopio electrónico.

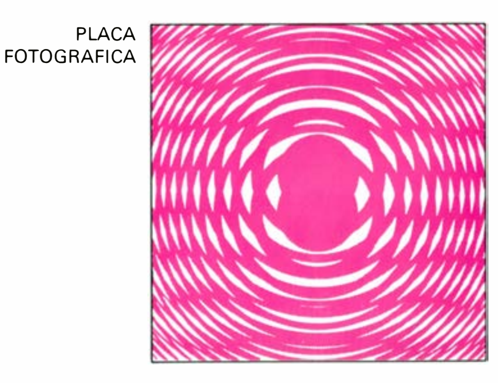
ocurrió que si impresionábamos los datos del radar sobre una película fotográfica o un medio similar, e iluminábamos las impresiones resultantes con un haz de luz coherente, regeneraríamos réplicas ópticas en miniatura de microondas que hubieran incidido originalmente sobre la abertura de recepción del sistema de radar. Asimismo, podríamos obtener imágenes de microondas de una manera sencilla y adecuada. Partiendo de esta idea, desarrollamos una teoría de la holografía bastante completa, paralelamente, en muchos aspectos, al trabajo anterior de Gabor que no conocíamos entonces. Nuestra teoría difería de la de Gabor en varios puntos. Introducía por primera vez la idea de la holografía fuera del eje, con objeto de evitar el problema de la doble imagen, e introducía, además, el empleo de una red de difracción en combinación con una placa de zonas de Fresnel para corregir la dispersión lateral y longitudinal, permitiendo en consecuencia operar con luz blanca. Así pues, aunque la holografía fuera del eje se identifique a menudo con el láser, esta forma de holografía se desarrolló en realidad en un contexto de luz completamente blanca.

La compensación de la placa de zonas tiene, por supuesto, una limitación importante. La dispersión longitudinal se corrige sólo para un punto imagen sencillo, concretamente aquel punto cuya componente del holograma se alinea con la placa de zonas de Fresnel compensadora. Para holografía de exhibición esta limitación es un inconveniente,



LA ESCALA DE LA IMAGEN DE DIFRACCION en ojo de buey formada por un objeto puntual es proporcional a la longitud de onda de la luz incidente. En la convención adoptada en los diagramas expuestos es esta página se representa luz de dos longitudes de onda diferentes mediante frentes de onda trazados a dos intervalos diferentes usando líneas de color de dos grosores distintos. El diagrama de la parte superior pone de manifiesto que el modelo formado por la luz de longitud de onda más larga, o luz más roja, es mayor que el formado por luz de longitud de onda más corta, o luz más azul. La imagen formada por un haz de luz compuesto de una amplia banda de longitudes de onda resultaría difuminada en los bordes; asimismo, el holograma constituido aparecería borroso, aunque se le observara luz perfectamente monocromática. El diagrama inferior muestra un holograma hecho con luz perfectamente monocromática, pero visto con luz policromática. El holograma actúa como una lente, dando lugar a que parte de la luz incidente se convierta en una onda divergente que vuelve a proyectar una imagen virtual, cuya resolución es, en teoría, la misma que la del objeto original. Sin embargo, la distancia entre el holograma y la imagen formada para cada longitud de onda es proporcional a la longitud de onda. De aquí que, para la luz blanca, que consiste en todas las longitudes de onda visibles, no exista ningún plano donde se pueda focalizar toda la luz y la imagen resultante sea borrosa *(dispersión longitudinal).*





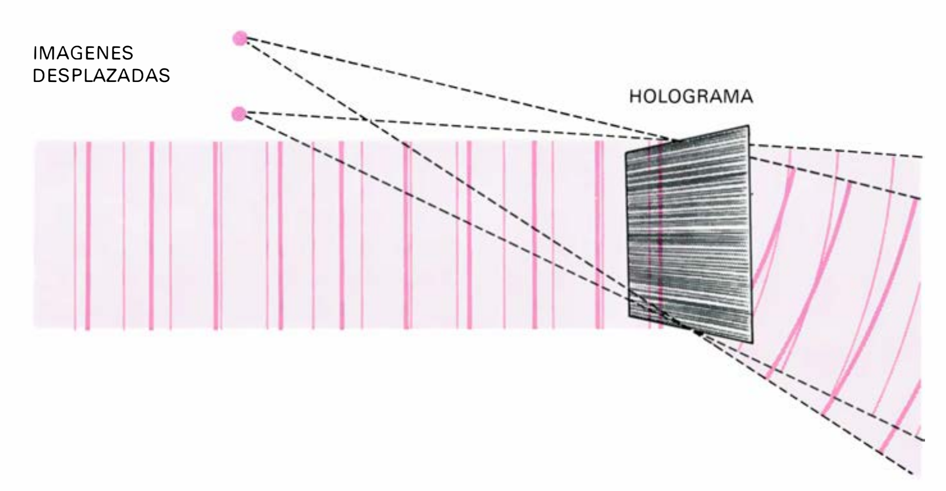
LA FUENTE DE LUZ EXTENSA es como un conjunto de fuentes puntuales, cada una de las cuales ilumina un objeto puntual desde una dirección ligeramente distinta. Las imágenes en ojo de buey se proyectan en direcciones distintas, resultando desplazadas en el holograma. Así, una fuente especialmente incoherente, o extensa, produce una imagen borrosa, donde las franjas de interferencia más finas desaparecen primero.

puesto que lo que se desea ver generalmente es la imagen completa y no sólo una pequeña parte de la misma. A la hora de aplicarlo al radar, este defecto no suponía un gran problema ya que la imagen se va grabando punto a punto mientras se pasa el registro de los datos del radar por una abertura, poniendo así uno tras otro cada punto imagen en posición para corregir su dispersión.

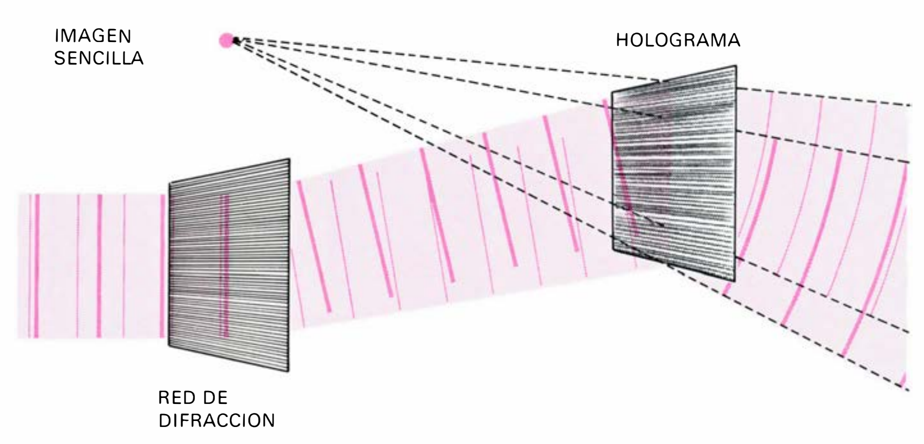
Este capítulo de la historia de la holografía se cerró sin apenas resonancia. En 1958, cuando estuvo completo nuestro sistema experimental de radar aerotransportado, pudimos verificar tales teorías y otras muchas que nuestro grupo había formulado. Sin embargo, desde el punto de vista de la ingeniería del diseño, preferimos la fuente de arco de mercurio a la fuente de luz blanca, y como nuestro sistema utilizaba razones de compresión del orden de sólo 100, encontramos que no se necesitaba corrección para la dispersión lateral ni longitudinal. La lámpara de arco de mercurio resultaba completamente ajustada e hicimos con ella reconstrucciones de alta calidad durante varios años, hasta 1962, en que cambiamos a fuentes de laser.

En 1963 se introdujo una nueva mejora en la holografía, al decidirnos a corregir otro defecto en el proceso holográfico. Con luz coherente se forman imágenes de difracción desde cualquier centro de dispersión que se encuentre en el recorrido de la luz. Así, el polvo sobre las superficies ópticas, las burbujas y otras oclusiones en las lentes añaden su imagen de difracción al holograma, que a su vez los transmite a la imagen. Introdujimos un difusor entre el objeto y la fuente de luz con objeto de eliminar estas imágenes. La técnica dio resultado y nos permitió formar imágenes de una calidad comparable a las de la fotografía convencional. Sin embargo, como el difusor tiende a difundir la luz dispersada por el objeto en ángulos más abiertos, las relaciones de expansión son significativamente mayores con esta técnica, incrementándose así las exigencias de coherencia, pero no más allá del alcance de la lámpara de arco de mercurio.

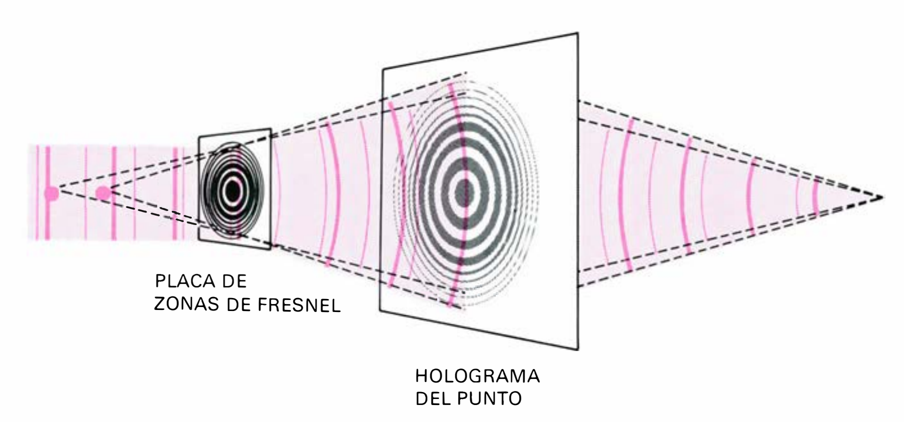
Nuestro siguiente paso fue construir hologramas de objetos reflectores tridimensionales, en vez de las transparencias que habían constituido hasta entonces el objeto de la holografía. Hacer un holograma de un objeto tridimensional es mucho más difícil por varias razones; una de las cuales es una exigencia de coherencia temporal grandemente potenciada. Esto puede entenderse rápidamente si se piensa en la exigencia de



EL HOLOGRAMA FUERA DEL EJE, inventado por el autor y su colaborador Juris Upatnieks en 1960, se comporta como una red de difracción de rectas paralelas, que desvía la luz en un ángulo proporcional a la longitud de onda. Las imágenes de longitudes de onda diferentes se desplazan lateralmente; la imagen de longitud de onda mayor, se desplaza más.



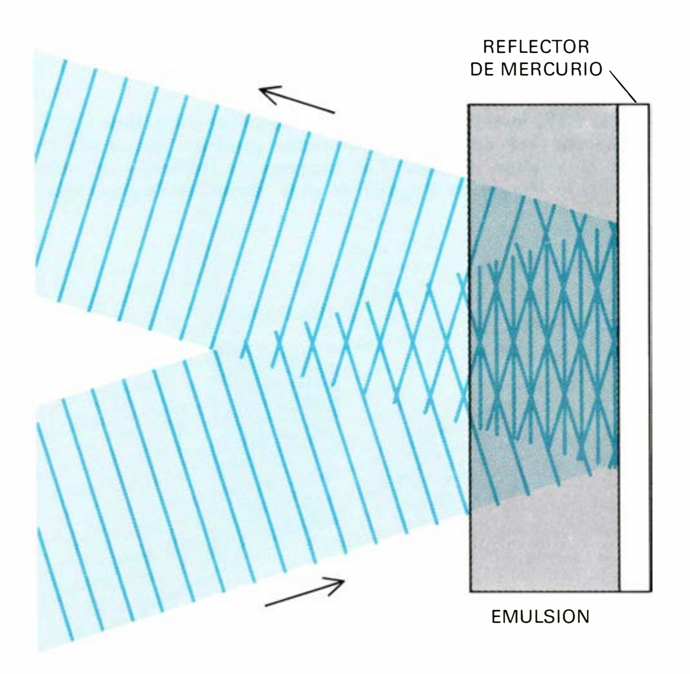
LA DISPERSION LATERAL SE CORRIGE mediante la inserción de una red de difracción en el camino del haz de luz incidente. La red se escoge cuidadosamente de modo que tenga un espaciado igual al espaciado medio de las franjas del holograma. Por tanto, puede usarse para desviar la luz contrariamente al modo en que lo hace el holograma, compensando exactamente la dispersión lateral originada por éste. En resumen, las imágenes formadas por dos longitudes de ondas diferentes se enfocan en el mismo punto, que es lo que se pretendía conseguir.

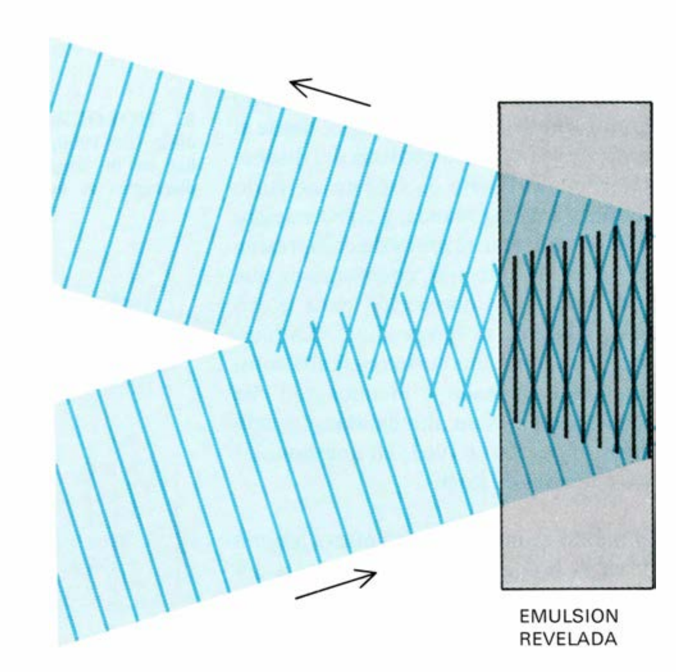


LA DISPERSION LONGITUDINAL SE CORRIGE mediante la inserción de una placa de zonas de Fresnel en el camino del haz de luz incidente. El dispositivo predispersa longitudinalmente un haz de luz blanca, formando una imagen virtual un tanto desplazada para cada longitud de onda, siendo las longitudes de onda más cortas las que se enfocan más lejos desde la placa de zonas. Esta colección de imágenes desplazadas sirve a su vez como fuente de luz para el holograma (en este caso también un holograma de un solo punto). El holograma que actúa exactamente igual que una placa de zonas desvía más la luz que tiene mayor longitud de onda reuniendo todos los colores en un punto común. La combinación de red de difracción de rectas paralelas y placa de Fresnel reducen las exigencias de coherencia para construir y visualizar hologramas fuera del eje.

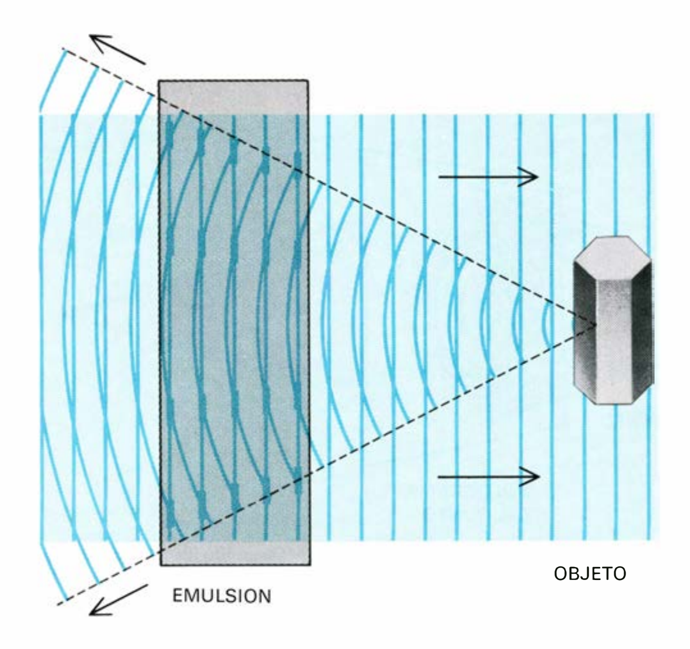
coherencia temporal entendida como longitud de coherencia. Al preparar nuestro dispositivo, dividimos un rayo láser en dos partes. Una, que servía de rayo de referencia, se dirigía a la placa fotográfica mediante un espejo; la otra parte del rayo era reflejada desde el objeto a la placa fotográfica. Para que estos dos rayos entraran en interferencia, los caminos que recorrieran no debían diferir, en cuanto a su longitud, en más de la longitud de coherencia de la luz. Por tanto, medimos los caminos y los ajustamos de modo que fueran aproximadamente iguales. Como la longitud de coherencia de la luz láser mide generalmente muchos centímetros, pudimos contar con el factor imprecisión. Si hubiéramos estado usando una lámpara de arco de mercurio, cuya longitud de coherencia es normalmente menor de 0.1 milímetros, hubiéramos tenido que igualar los caminos con mucho mayor justeza.

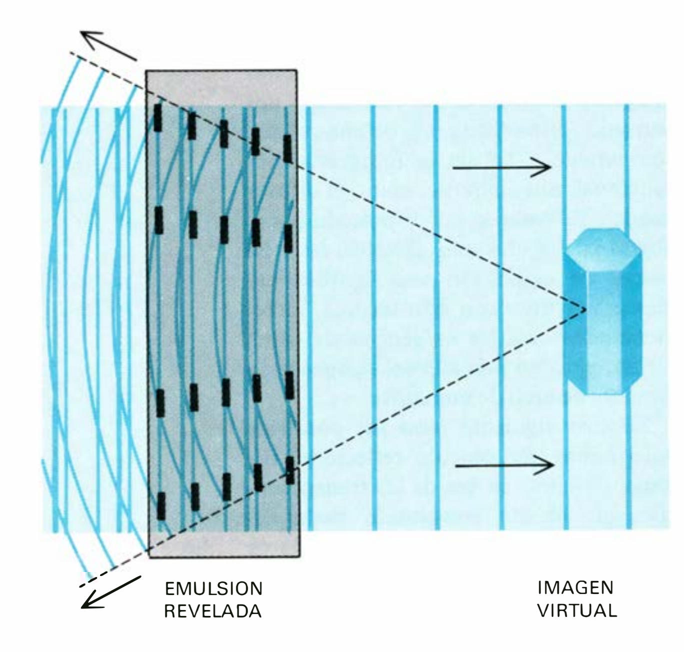
Sin embargo, suponiendo que el objeto tenga profundidad, la longitud del haz objeto dependerá de la situación de un putno en el objeto, y no podríamos igualarla entonces para todos los puntos objeto. Precisaríamos en ese caso una fuente de luz con una longitud de coherencia suficientemente grande para abarcar la profundidad del objeto. Si la profundidad del objeto es, por ejemplo, de medio metro, la longitud de coherencia debería ser de un metro por lo menos,





EL PROCESO DE LIPPMANN para fotografiar en color, inventado por el físico francés Gabriel Lippmann en 1891, produce franjas de interferencia en una emulsión fotográfica gruesa, al permitir que la luz incidente vuelva a reflejarse con la ayuda de un baño de mercurio en la parte posterior de la emulsión *(diagrama de la izquierda).* **Los depósitos de plata que se forman en las superficies de brillo máximo se fijan luego por revelación de la placa. En el proceso de lectura** *(diagrama de la derecha).* **La luz incidente se refleja parcialmente en las capas sucesivas de los granos de plata de la emulsión. Las reflexiones se van reforzando a la longitud de onda de la luz usada para la exposición de laplaca.**





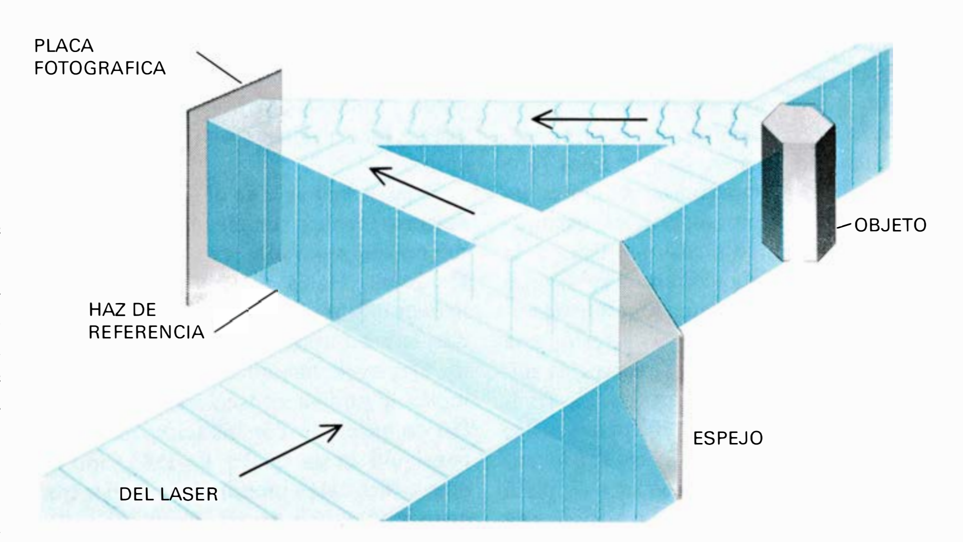
EL METODO DENISYUK para holografía, que consiste fundamentalmente en una combinación de los procesos holográficos de Lippmann y Gabor, funciona haciendo pasar un haz de luz coherente por una emulsión fotográfica, reflejando la luz de un objeto e impresionando la interferencia producida en la emulsión entre el haz incidente y el haz reflejado *(izquierda).* **El holograma se observa al dirigir luz blanca hacia el mismo, que selecciona por refuerzo de la luz de sólo una longitud de onda, formando una imagen tridimensional del objeto** *(derecha).*

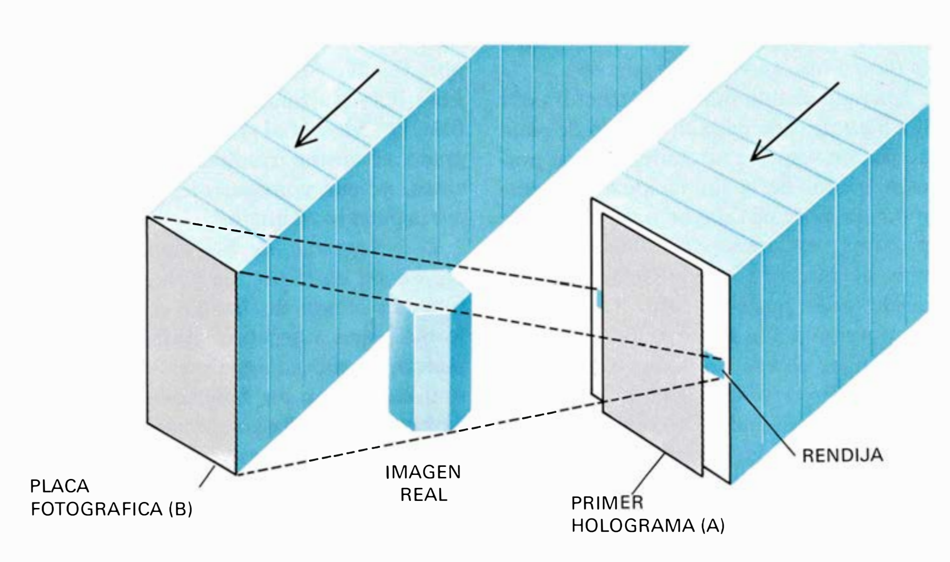
puesto que tal sería la diferencia de la longitud del camino de la luz entre un rayo de luz reflejado desde el frente del objeto y otro reflejado desde la parte más distante al mismo. La longitud de coherencia correspondiente que se exige para la construcción de hologramas de transparencias sería, en general, de unos 0,01 milímetros. Esta nueva exigencia de coherencia es independiente de la descrita anteriormente en conexión con las transparencias, pero es tan grande que la exigencia anterior resulta comparativamente despreciable. En esta etapa de nuestro trabajo hubo que recurrir, pues, al láser que, afortunadamente, había ido desarrollándose durante el mismo período y que apareció en un momento especialmente oportuno para nuestros intereses.

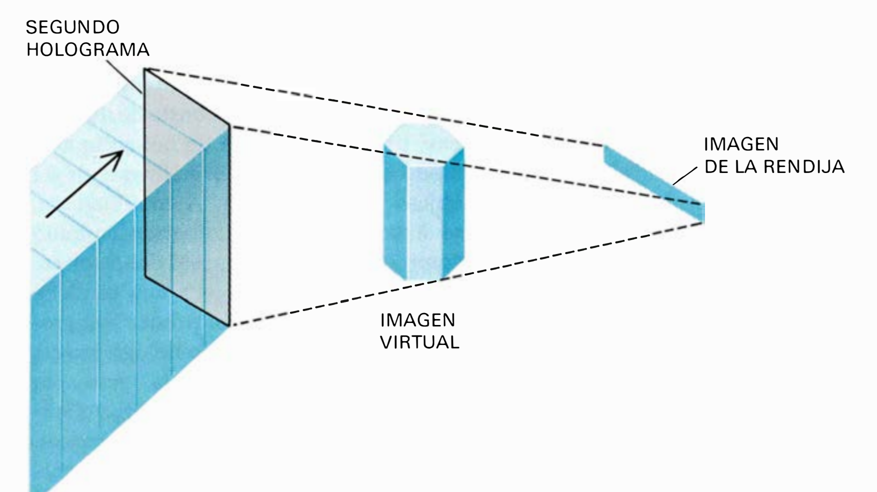
Hasta ese momento, los procesos de construcción y visualización de hologramas habían tenido unas exigencias de coherencia similares; hecho que salta a la vista si consideramos que la observación es, simplemente, una reconstrucción de los rayos que habían intervenido en el proceso de formación. En este proceso de visualización, sin embargo, no existe la contrapartida del fenómeno de grandes caminos ópticos producidos en las reflexiones de las diferentes profundidades del objeto, puesto que el proceso no implica ninguna reflexión de este tipo. De aquí que esta nueva exigencia de coherencia no afecte al proceso de visualización. Por consiguiente, los hologramas de objetos tridimensionales que requieren un láser para su formación pueden observarse con luz de una lámpara de arco de mercurio, que es el modo habitual.

Contemporáneamente, en 1962, el investigador ruso Yu. N. Denisyuk publicó un importante avance, en el que el proceso de la holografía se combinaba con una forma de la fotografía en color inventada en 1891 por el físico francés Gabriel Lippmann. El holograma de Denisyuk podía producir imagenes monocromáticas o de color cuando se visualizaba con luz blanca desde una fuente puntual.

En el proceso de Lippmann se forma una imagen sobre una placa fotográfica de alta resolución. La placa se refuerza en el lado de la emulsión por un baño de mercurio, de modo que la luz que entra en la emulsión desde el lado de la placa se refleja y atraviesa la emulsión por segunda vez. Los dos rayos de luz que viajan en direcciones opuestas interfieren inmediatamente, incluso si la fuente de luz tiene un amplio espectro de longitudes de onda, puesto que la diferencia de







LOS HOLOGRAMAS EN ARCO IRIS SE REALIZAN en dos etapas. Primero se crea un holograma "patrón" convencional *(arriba)*. **A continuación, se sitúa sobre el holograma una rendija horizontal y se forma una imagen real, o frente de onda convergente, iluminando el holograma con la versión conjugada, o de tiempo invertido, del haz de referencia** *(centro)*. **Esta imagen sirve luego como objeto para un segundo holograma que se impresiona cerca del espacio de la imagen real con ayuda de un haz de referencia convergente, el cual se inclina verticalmente. Cuando se ilumina el segundo holograma con una fuente de luz blanca, cada longitud de onda forma una imagen de la rendija en una posición vertical diferente** *(abajo)*. **Cuando el observador mira a través de cualquiera de estas imágenes de la rendija, la escena original se percibe en tres dimensiones, dotada de paralaje horizontal, en un solo color. Al subir y bajar la cabeza el observador, la imagen aparece sucesivamente en los diferentes colores del arco iris, aunque carece de paralaje vertical. La técnica aquí mostrada fue inventada por Benton, de la Polaroid Corporation.**

camino es casi cero. La fuente de luz también puede ser incoherente espacialmente, puesto que los mismos rayos, o rayos muy adyacentes, se superponen. Así pues, se forman franjas finas (que son realmente superficies brillantes interespaciadas con superficies oscuras) alineadas de modo casi paralelo a la emulsión. La distancia entre las franjas es sólo de la mitad de una longitud de onda de la luz. Por tanto, a través de una emulsión con un grosor típico de 15 micras se forman unas 30 franjas. Las franjas producen depósitos de plata en las regiones brillantes como en el proceso fotográfico normal. Los granos de plata depositados en cada lugar tienen una concentración relacionada con el brillo de la imagen en esa región; más aún, su espaciado está relacionado con el color de la imagen en esa posición *(véase la ilustración superior de la pág. 66)*.

Cuando se ilumina esta estructura con luz blanca, cada capa de granos de plata actúa de espejo, al reflejar una porción menor de la luz incidente y transmitir la porción mayor a la superficie próxima, en donde se refleja un poco más de luz. El total de la luz reflejada desde cada posición sobre la placa es proporcional a la luz original que formó las franjas en esa posición. Así pues, el observador ve sobre la placa una imagen del objeto original tal y como ocurre en la fotografía convencional. Y lo que es más importante, la imagen es a todo color debido a los efectos de interferencia entre las distintas capas. Cada capa refleja luz de todos los colores, pero hay un refuerzo de sólo ciertos colores en razón de la distancia entre las franjas. Puesto que el espaciado de las franjas vino determinado por la luz original, los colores que resultan reflejados son los correspondientes a los de la luz original.

El trabajo de Denisyuk, una de las piedras angulares de la holografía, combinaba el trabajo de Lippmann y Gabor al usar luz coherente e impresionando, en vez de una imagen, las imágenes de difracción producidas por el objeto. En el método de Denisyuk, la luz pasa a través de la placa, incide sobre un objeto, y vuelve reflejada a la placa atravesándola por segunda vez pero en dirección opuesta. Igual que en el proceso de Lippmann, se forman franjas en toda la profundidad de la emulsión. Resulta así un holograma que tiene las características de la placa de Lippmann y puede verse con una fuente de luz blanca de extensión espacial limitada, tal como una lámpara de proyector o una bombilla incandescente ordinaria *(véase la ilustración inferior de la pág. 66)*. Más aún, usando tres rayos de luz coherente a tres longitudes de onda discretas, una por cada color primario, puede hacerse un holograma que produzca una imagen a todo color.

El paso siguiente se daría al combinar nuestra técnica, empleando un haz de referencia separado, con la técnica de Denisyuk. Hacia principios de 1964, se habían difundido ampliamente los informes sobre nuestra técnica y los trabajos de Denisyuk habían sido traducidos al inglés, y podían conseguirse fácilmente. Sin embargo, la combinación real no se consiguió hasta finales de 1965, año en que se hizo casi simultáneamente por tres grupos. Nuestro grupo comunicó y demostró tales hologramas en la reunión de la Sociedad Americana de Optica celebrada en la primavera de 1966; lo mismo hizo en Michigan otro grupo dirigido por G. W. Stroke y A. E. Labeyrie. Pero Nile Hartman, que trabajaba en el Battelle Memorial Institute, había logrado el mismo resultado varios meses antes, y, por consiguiente, se le concedió la patente en litigio.

El holograma de luz blanca de Denisyuk es, en efecto, un holograma y un filtro de color de banda estrecha que forman una estructura única. Por supuesto, puede conseguirse la misma visualización con luz blanca con un filtro de banda estrecha combinado con otros tipos de hologramas, de suerte que la principal ventaja de este tipo de hologramas reside en evitar la inconveniencia de un filtro de color separado.

En 1969 Stephen Benton, de la Polaroid Corporation, comunicaba otro avance importante en la holografía de exhibición. Su método consta de dos etapas; primero se forma un holograma de modo convencional (normalmente de un objeto tridimensional). A continuación, se hace un segundo holograma usando como objeto la imagen real del primero. En este proceso se coloca una estrecha rendija horizontal en el primer holograma. Como cada punto de holograma reproduce la imagen completa, pero sólo desde un punto de vista, la rendija elimina todo el paralaje en el plano vertical. (El paralaje, elemento esencial en la percepción de la tridimensionalidad, es el desplazamiento aparente de un objeto cuando se ve desde dos puntos diferentes.) El segundo holograma, con iluminación coherente, forma una imagen (sea virtual o real) del objeto original, así como una imagen real de la rendija. Para ver la imagen completa, el observador debe colocar su ojo en esta imagen de la rendija *(véase la ilustración de la pág. 67)*. Si mueve la cabeza verticalmente, la imagen desaparece; así pues, no hay paralaje vertical. Sin embargo, la imagen puede resultar bastante brillante; toda la luz que normalmente se difractaría a una distancia vertical de un metro aproximadamente en la posición del observador, ahora se limita a la imagen de la rendija y la ganancia de luz puede ser varios centenares de veces mayor.

Hay, sin embargo, en este método una ventaja mucho más importante. Si se quisiera iluminar el holograma con luz de dos longitudes de onda diferentes, la dispersión lateral produciría dos imágenes de la rendija desplazadas verticalmente, una para cada color. El observador vería la imagen completa en un color, cuando colocara sus ojos en una de las imágenes de la rendija y, en otro color, cuando los colocara en la otra. Extrapolando este proceso a iluminación con luz blanca, donde la imagen de la rendija se difuminaría en un espectro continuo, se encuentra que, para ver el holograma, el observador puede colocar sus ojos en cualquier punto del espectro. Donde quiera que los coloque, verá la imagen nítida y clara en el color que corresponda a esa parte del espectro. Al mover la cabeza verticalmente la imagen percibida cambia de color pero permanece con la misma nitidez y claridad. Aunque la cabeza no está en el plano de la imagen de la rendija, seguirá viendo la imagen completa, pero el color cambiará de arriba abajo. Por lo cual tales hologramas han sido llamados, a veces, hologramas en arco iris.

A cambio de renunciar al paralaje vertical, sosteniendo el paralaje horizontal normal, se obtienen hologramas que pueden verse con luz blanca sin necesidad de ningún filtro de banda estrecha ni externo ni incorporado. Más aún, se utiliza toda la luz blanca con preferencia a sólo una banda estrecha del espectro, haciendo que los hologramas Benton resulten extraordinariamente brillantes. La pérdida del paralaje vertical no suele tener mayor importancia, puesto que el observador mueve normalmente la cabeza en sentido horizontal, casi nunca en vertical. Obviamente, si el observador moviera los ojos verticalmente, cada ojo recibiría la misma imagen y no habría percepción de la tridimensionalidad. En suma, el método Benton es un trueque favorable de aspectos de poca monta con otros importantes.

Otra mejora notoria, la técnica compuesta o múltiple, remonta algunas de las limitaciones básicas de las técnicas holográficas más convencionales, consi-

guiendo resultados impresionantes. En esta técnica, se sintetiza el holograma a partir de muchisimas fotografías convencionales obtenidas al modo usual. Las fotos, tomadas desde posiciones diferentes, constituyen muchas vistas del objeto y, en su conjunto, reúnen toda la información esencial contenida en un holograma. El método compuesto sintetiza las fotografías en un único holograma.

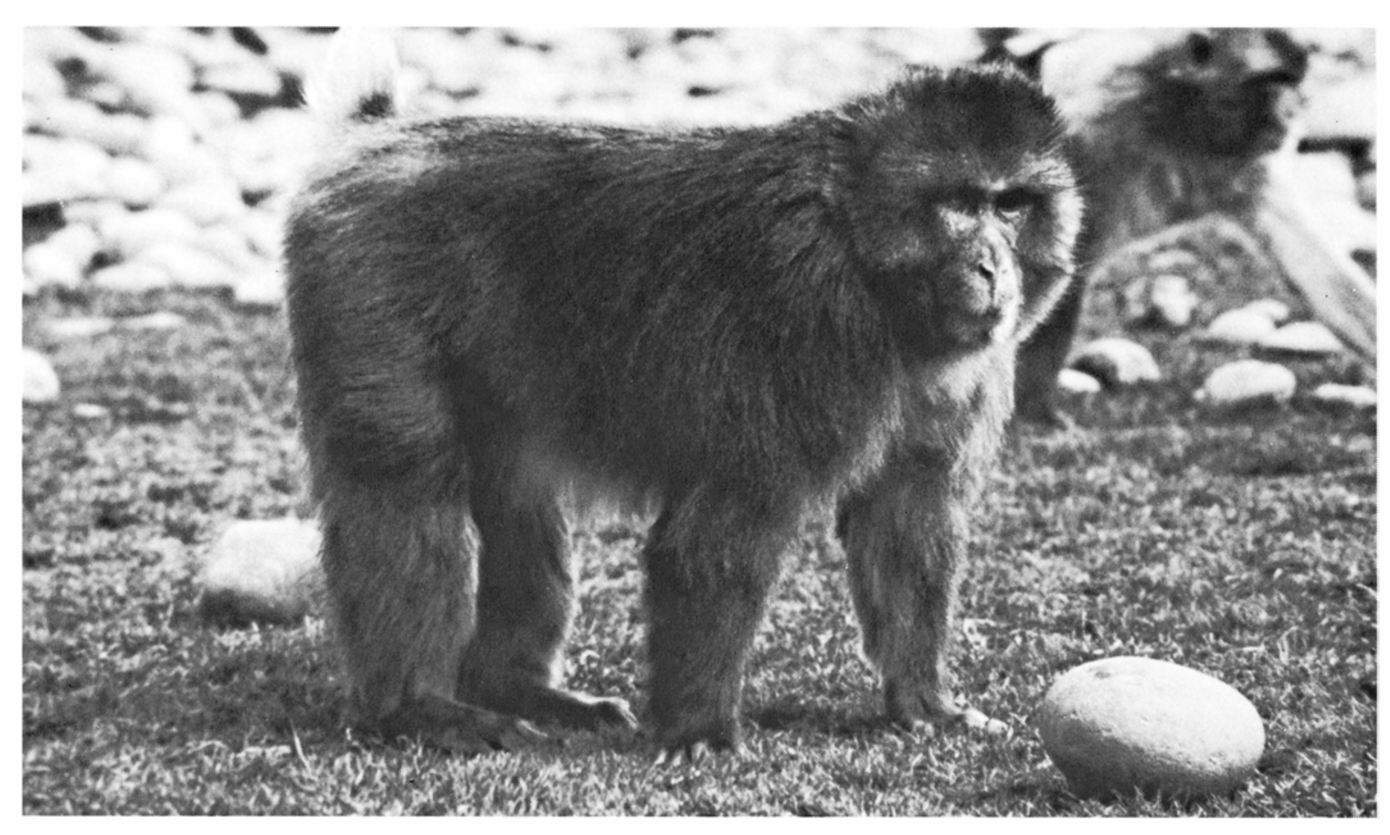
El primer informe sobre hologramas compuestos se publicó en 1967 por Robert V. Pole, de la International Business Corporation. En los años siguientes, se lograron muchos avances que culminaron en un sistema perfeccionado en los últimos cinco años por Lloyd Cross y colaboradores en la Multiplex Company, quienes, por su parte, introdujeron varias innovaciones oportunas. En el método ideado por Cross, el objeto se sitúa en una plataforma que gira lentamente mientras una cámara cinematográfica toma cientos de fotogramas exactamente igual que en el cine convencional. Después, estos fotogramas pueden convertirse en un holograma simple del siguiente modo: el primer fotograma puede ser convertido en imagen con luz coherente, en una pantalla de vidrio esmerilado. La luz transmitida por la pantalla se impresionaría, como un holograma, en forma de una cinta larga y delgada de un milímetro de anchura y 20 centímetros de longitud, aproximadamente en cometido de la pantalla estriba en dirigir la luz desde todas las partes de la imagen a la rendija impresionadora. El sistema real difiere de éste en que, para dirigir la luz a la rendija, se usa una lente cilíndrica grande en el plano de la imagen en lugar del difusor. La imagen del holograma resultante es muy superior. Después de impresionar, a modo de holograma, un fotograma, el siguiente se proyecta de forma similar, construyéndose un segundo holograma en una posición adyacente y contigua al primero; y así sucesivamente. Esto se hace a gran velocidad por medio de una rendija vertical situada delante del film virgen, que avanza por la anchura de una rendija, entre dos exposiciones.

El holograma compuesto resultante se coloca luego en un arco que corresponde al ángulo de rotación del objeto, y se ilumina como un holograma convencional. Cada holograma componente proyecta una imagen virtual sin profundidad o paralaje, exactamente igual a la imagen de la que se hizo el holograma. Sin embargo, cuando un observador contempla la película, cada ojo mira a un punto de la imagen a través de hologramas distintos. Las imágenes formadas en las dos retinas son ligeramente diferentes, como en un caleidoscopio, y, por consiguiente, se percibe una imagen tridimensional. Al mover el observador su cabeza lateralmente, se aprecian diferentes imágenes holográficas; así se producen las relaciones del paralaje de la vida real. Los numerosos fotogramas que contienen muchas vistas diferentes se han convertido, en efecto, en un holograma único. Los cambios de paralaje aparecen en pequeños saltos, por supuesto, pero no se notan más que los saltos en una secuencia de fotogramas de una película. En el plano vertical no hay paralaje, pero ello tampoco tiene ninguna trascendencia.

La técnica de Cross puede combinarse con la de Benton permitiendo así la visión con luz blanca. No se renuncia a nada esencial al hacerlo, puesto que, en cualquier caso, los hologramas compuestos carecen de paralaje vertical.

La técnica compuesta ha permitido construir hologramas de seres vivos sin recurrir a la iluminación con rayo láser. Más aún, puesto que cabe cierto movimiento del objeto, la técnica ha conseguido hologramas en los que puede verse al individuo moviendo la mano, levantándose o cambiando su expresión facial. Este movimiento se percibe viendo en secuencia las distintas partes del holograma, permitiendo que el observador recorra el holograma o quizás que éste gire.

Los hologramas compuestos se forman a partir de fotos hechas de modo convencional con luz blanca ordinaria. Cuando se combinan con la técnica de Benton, pueden verse también con luz blanca ordinaria. Puesto que el holograma se forma sólo de transparencias, y puesto que el cono de rayos que incide en cada parte del film es bastante estrecho, parece que el holograma podría construirse con la luz casi coherente de una lámpara de arco de mercurio o, aprovechando nuestra técnica original de *scanning,* con luz blanca ordinaria. En resumen, el proceso de Cross, que produce algunos de los hologramas más impresionantes y complicados, no requiere en absoluto del láser; por tanto, podía haberse usado incluso al final de la década de los cuarenta, cuando por primera vez Gabor ideó la holografía. De esta descripción de los métodos holográficos y sus exigencias de coherencia, destaca una observación fundamental: el proceso no depende para nada de una iluminación altamente coherente, ni en su proceso de realización ni en su proceso de visualización. En realidad, nunca fue dependiente.



MACHO ALFA O DOMINANTE de un grupo natural de macacos japoneses mantenido en condiciones seminaturales en un corral de unos 8000 m² de superficie, en el Oregon Regional Primate Research Center. Su nombre es Arrowhead o cabeza de flecha. El macho alfa dirige, defiende y controla al grupo. Arrowhead, que tiene 25 años, era ya quizá el macho alfa del grupo cuando éste se hallaba en libertad en Japón. Además de ser uno de los machos más bajos, carece de caninos y le falta un ojo; su posición dominante nunca ha sido amenazada.

HEMBRAS ADULTAS arrebujándose con sus hijos sobre un tocón de árbol en el corral del Oregón. Los hijos, generalmente únicos, nacen entre abril y mayo, permaneciendo junto a su madre durante todo el verano. Dos semanas después del nacimiento, la madre empieza a enseñarles a andar y reconocer los alimentos. Los macacos hembra protegen con gran celo a sus hijos hasta pasadas varias semanas.

Organización social de los macacos japoneses

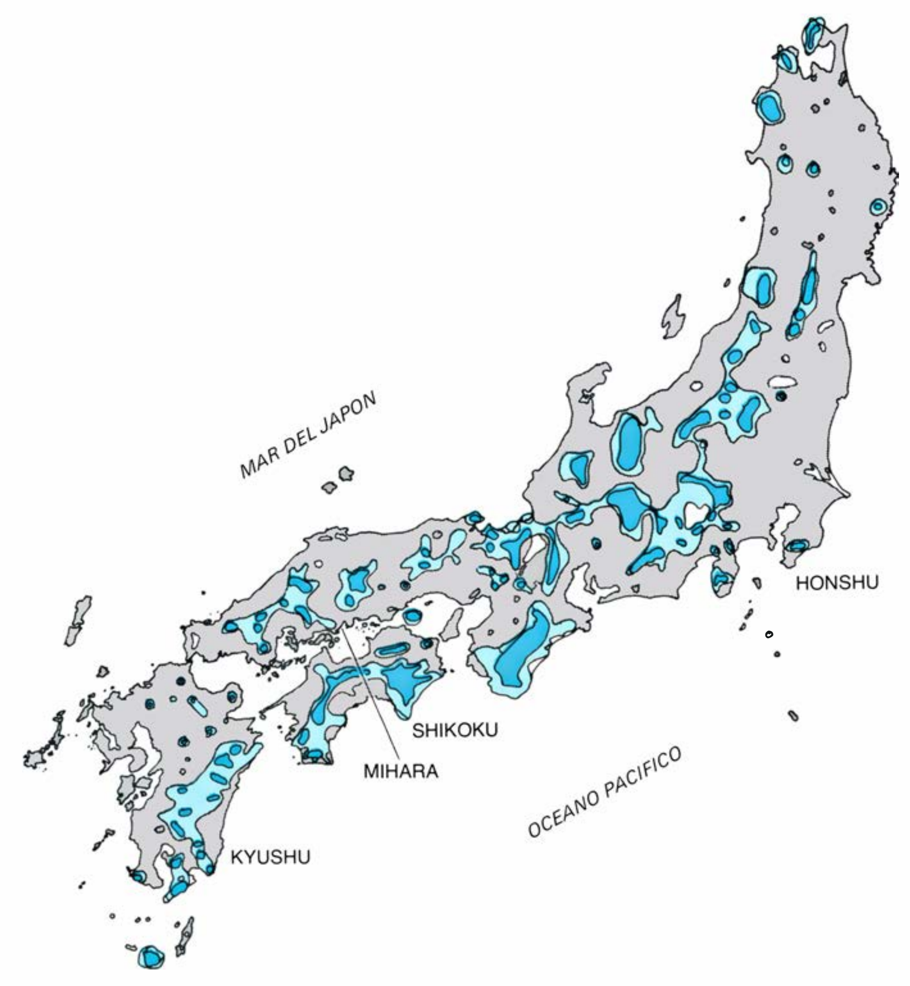
El estudio a largo plazo de un grupo natural de macacos japoneses confinados en un corral del Centro de Primatología de Oregón indica que las bases biológicas de su comportamiento se hallan ampliamente modificadas por los factores sociales

G. Gray Eaton

Los macacos japoneses son bien conocidos en su aspecto más turístico: los tres famosos monos de madera que ilustran el proverbio budista que reya "Ni hablar, ni oír ni ver", y que pueden encontrarse en cualquier tienda de "souvenirs". La elección de este modelo resulta adecuada, ya que los macacos son primates de extraordinaria inteligencia, emparentados con los macacos rhesus y con los monos de Gibraltar, que viven en grupos de gran cohesión formados por 50 a 150 individuos. Estos animales poseen una compleja organización social que, al igual que cualquier otra organización de este tipo, se ha producido como resultado de la interacción de una serie de factores biológicos y sociales. Lo que más llama la atención en los macacos japoneses es que el componente social de su organización presenta una influencia superior a la que sería de esperar.

En su hábitat natural en los bosques de Japón, los macacos son animales tímidos cuya observación resulta difícil, por lo que los etólogos interesados en estudiar sus formas de comportamiento se han visto obligados a suministrarles alimentos o a estudiarlos en cautividad. Durante los últimos seis años, y junto con mis colaboradores, hemos observado un grupo típico de macacos confinados en un corral de unos 8000 m² situado en el Oregon Regional Primate Research Center por espacio de más de 8000 horas. El grupo estudiado por nosotros procede de las cercanías de la ciudad de Mihara, en la prefectura de Hiroshima. Como resultado de las protestas de los agricultores de la región, el Japan Monkey Center capturó a la casi totalidad de los 49 miembros del grupo en 1964. Al año siguiente, los animales fueron trasladados al corral del Centro de Oregón, donde se encuentran desde dicha fecha.

Los macacos de nuestro grupo no habían sido estudiados con anterioridad, lo que hace que resulte difícil valorar el efecto que la falta de libertad pueda haber tenido sobre su comportamiento. Partimos, sin embargo, de la base de que probablemente el confinamiento y la provisión de alimentos aumenten la frecuencia de todo tipo de interacción social, ya que, al no verse obligados a ocu-



LA SUPERPOSICION DE LOS MAPAS realizados por K. Hasebe en 1932 *(color claro)* **y por K. Kishida en 1953** *(color oscuro)* **pone de manifiesto la disminución progresiva de las zonas de distribución de los macacos japoneses. Aunque los extremos de distribución, que son los que más al norte se encuentran de todos los primates existentes en la actualidad, no han variado en los últimos cincuenta años, la eliminación de los bosques por el hombre ha reducido su extensión. Sin embargo, en la actualidad, estos macacos son considerados como parte del patrimonio nacional, y su exportación se halla rigurosamente prohibida por el gobierno japonés.**

par largas horas en busca de alimentos, los animales disponen de más tiempo para el ocio. En algunos casos en que, en lugar de administrarles su ración habitual de alimento preparado, hemos esparcido cereales en el corral, los animales se han dedicado intensamente a la búsqueda del alimento, abandonando otras actividades.

El aspecto más interesante del comportamiento del grupo consiste en el dominio que unos pocos machos ejercen sobre el resto de los animales. A continuación, llama la atención el que las hembras viejas ataquen a hembras más jóvenes sin que éstas respondan a la agresión, y el que ciertas hembras adultas ataquen y pongan en huida a determinados machos. Pronto se pone de manifiesto la existencia, entre los macacos, de una rígida jerarquía de dominancia que recuerda a la existente en un gallinero. El puesto de mayor autoridad, correspondiente al líder o macho "alfa" se halla casi siempre ocupado por un macho adulto, que en ocasiones no alcanza este rango hasta cumplidos los 18 o 19 años de edad. (Los machos llegan a la pubertad a los cuatro años, y completan su crecimiento entre los ocho y los diez años.) A continuación se encuentran cinco o seis machos "subdominantes", y, por debajo de ellos, las hembras adultas, que alcanzan la pubertad a los tres años y completan su crecimiento entre los seis y los diez años; estas hembras, junto con sus des-



CORRAL CUBIERTO DE HIERBA, en el Oregon Regional Primate Research Center, donde se encuentran los 230 animales que forman parte del grupo de macacos japoneses estudiado por el autor. (Los monos son los puntos negros que se observan en el recinto.) Las paredes del corral están formadas por planchas de acero acanalado de 3,5 metros de altura, inclinadas hacia dentro en un ángulo de 15 grados. En el recinto se encuentran tocones y postes de madera para que los animales, que viven generalmente en el suelo pero pueden comportarse como arbóreos, puedan trepar; los caminos que se observan en la hierba fueron abiertos por los propios animales al desplazarse entre sus zonas predilectas. Véanse tres torres de observación; a la izquierda se ve parte de un nuevo corral para un estudio de población.

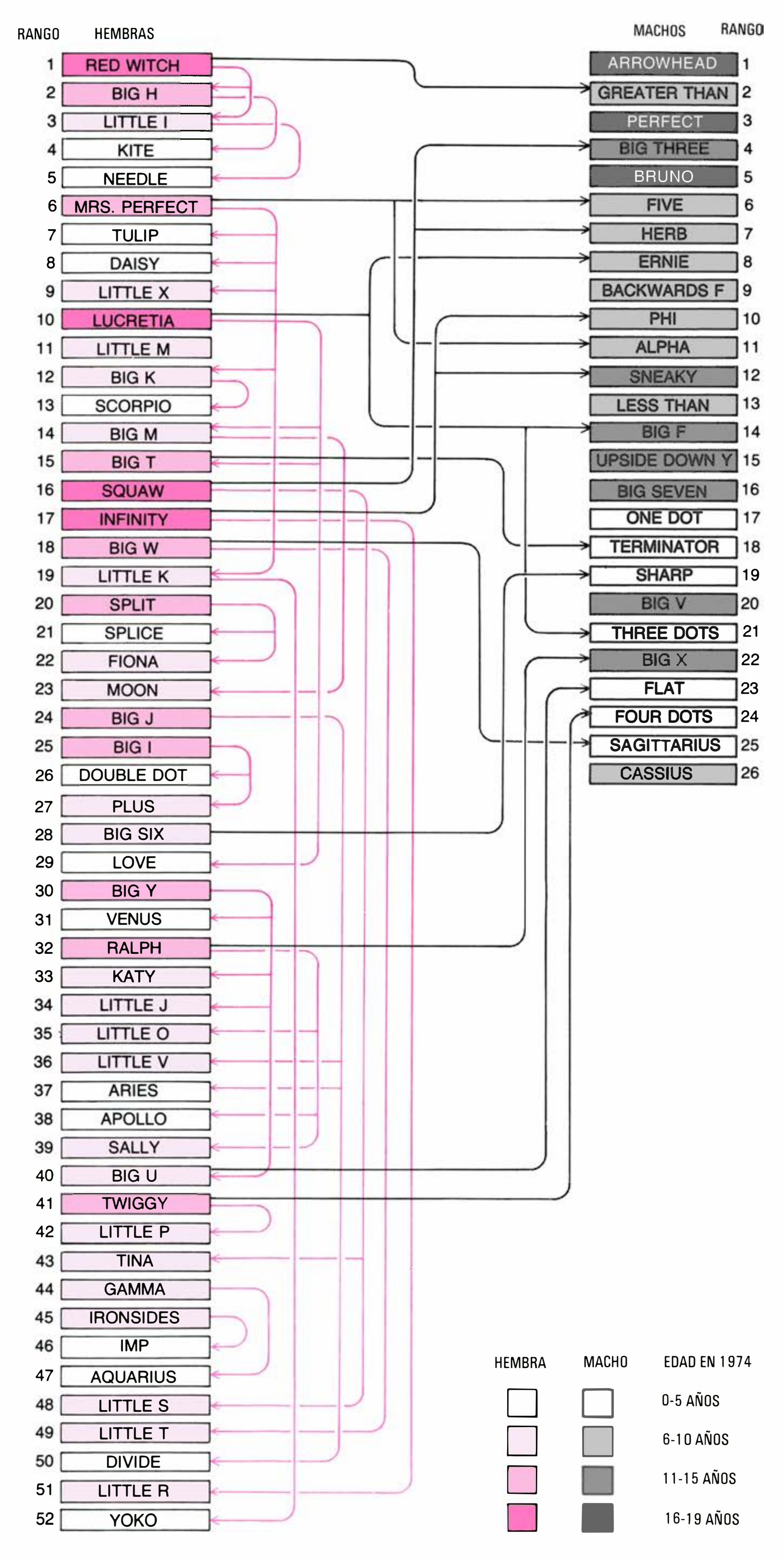
cendientes infantiles y juveniles, constituyen el estrato medio de la jerarquía. Los machos restantes constituyen el último estrato jerárquico, y, en condiciones naturales, viven en la periferia del grupo.

El rango de un animal no se halla en correlación con su agresividad, lo que hace imposible el determinar la jerarquía a partir del número de ataques. De hecho, el grupo se halla organizado de forma tal que, cuanto más elevado es el rango de un individuo, menor es el número de ataques a que se halla expuesto, y viceversa. El rango de dominancia es básicamente lineal, aunque en algunas ocasiones se produzcan inversiones del mismo, como cuando el animal X persigue a *Y*, *Y* persigue a *Z* y, finalmente, *Z* persigue a *X*.

En consecuencia, los animales de rango elevado no son siempre los más agresivos. En nuestro grupo, los macacos situados en segundo y tercer lugar, es decir, Greater Than y Perfect, son más agresivos que el macho alfa, Arrowhead, entre otras razones porque Arrowhead es más . respetado” por otros miembros del grupo y, en consecuencia, no se ve obligado a afirmar constantemente su puesto atacando o amenazando a otros animales del mismo. Así, por ejemplo, cuando los animales son alimentados, Arrowhead dispone de más espacio para comer que Greater Than, y en caso de que se ofrezca al grupo un alimento especialmente atractivo, como un cacahuete, la mera presencia de Arrowhead basta para mantener a otros animales a distancia, mientras que Greater Than tiene que echarles de sus cercanías con frecuencia.

Paradójicamente, los animales dominantes no se sitúan en la cumbre de la jerarquía como resultado de su capacidad para la lucha ni de sus características físicas, como el tamaño. Arrowhead es uno de los machos adultos de menor tamaño, carece de caninos y ha perdido un ojo. Sin embargo, su puesto de macho alfa no se ha visto nunca amenazado. Además, en los raros casos en que ataca a otro macho, éste no intenta pelear, sino que se limita a huir. Entre los machos subdominantes, Bruno y Perfect carecen también de caninos, y Greater Than alcanzó el rango que posee en la actualidad antes de que sus caninos hubiesen crecido totalmente. En consecuencia, si ni el tamaño del animal ni los caninos determinan que el animal alcance o mantenga su rango, ¿en qué se basa la jerarquía de los macacos?

Después de prolongadas observaciones realizadas en macacos japoneses en



JERARQUIA DE DOMINANCIA ENTRE LOS ADULTOS del grupo de macacos de Oregón en 1974. Las jerarquías correspondientes a las hembras *(gris)* y a los machos *(color)* se presentan por separado, dados los cambios que comporta la reproducción, momento en que los machos subordinados atacan a las hembras. Las flechas corresponden a las relaciones materno-filiales, y ponen de manifiesto la influencia que la jerarquía de la madre posee sobre el rango del hijo.

condiciones naturales, los etólogos han llegado a la conclusión de que el rango de un animal depende en gran parte del rango de su madre. El papel de la madre en la fijación del rango de sus hijos se pone de manifiesto ya en las peleas entre juveniles, que se producen generalmente como resultado del endurecimiento del juego. Uno de los animales juveniles chilla, su madre acude en su ayuda y muerde a su rival, y ello provoca la intervención de la madre del animal agredido. A continuación, las madres pelean, hasta que la dominante persigue, junto con su hijo, a la madre subordinada, cuyo hijo o hija escapa con ella. Cuando el episodio se ha repetido varias veces, el hijo de la madre subordinada huye del hijo de la madre dominante aún cuando ésta no se halle en las cercanías. En algunos raros casos, un macaco puede resultar excepcionalmente agresivo, y alcanzar un rango superior al que posee su madre, pero este tipo de situación es excepcional.

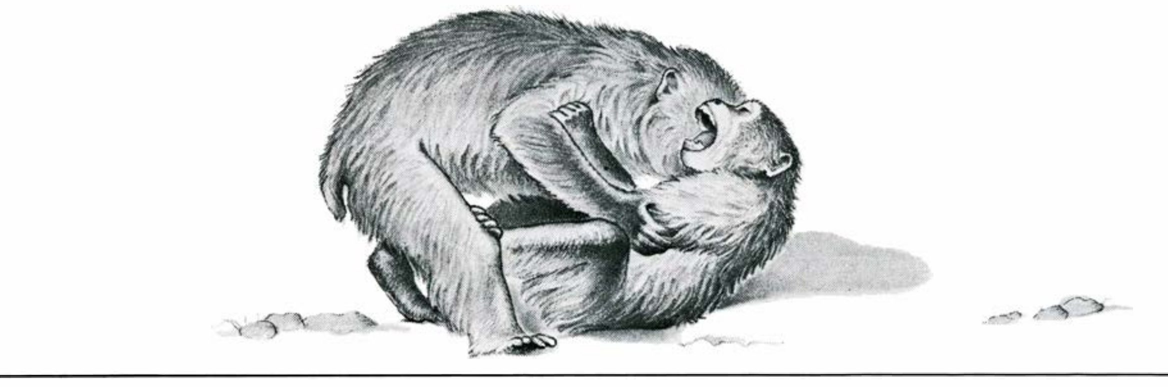
En nuestro grupo existen varios machos adultos que son aún defendidos por sus madres. Cuando el macho subdominante Greater Than, que ocupa actualmente el segundo lugar de la jerarquía, se encontraba aún en tercer lugar, era a veces atacado por el segundo macho, One-Eye, y por el macho alfa, Arrowhead. En estos casos, su madre, Red Witch, distraía a los agresores saltando sobre su lomo, permitiendo así que su hijo escapase.

En estado natural, los hijos de las hembras de mayor rango permanecen en el

1



2



3

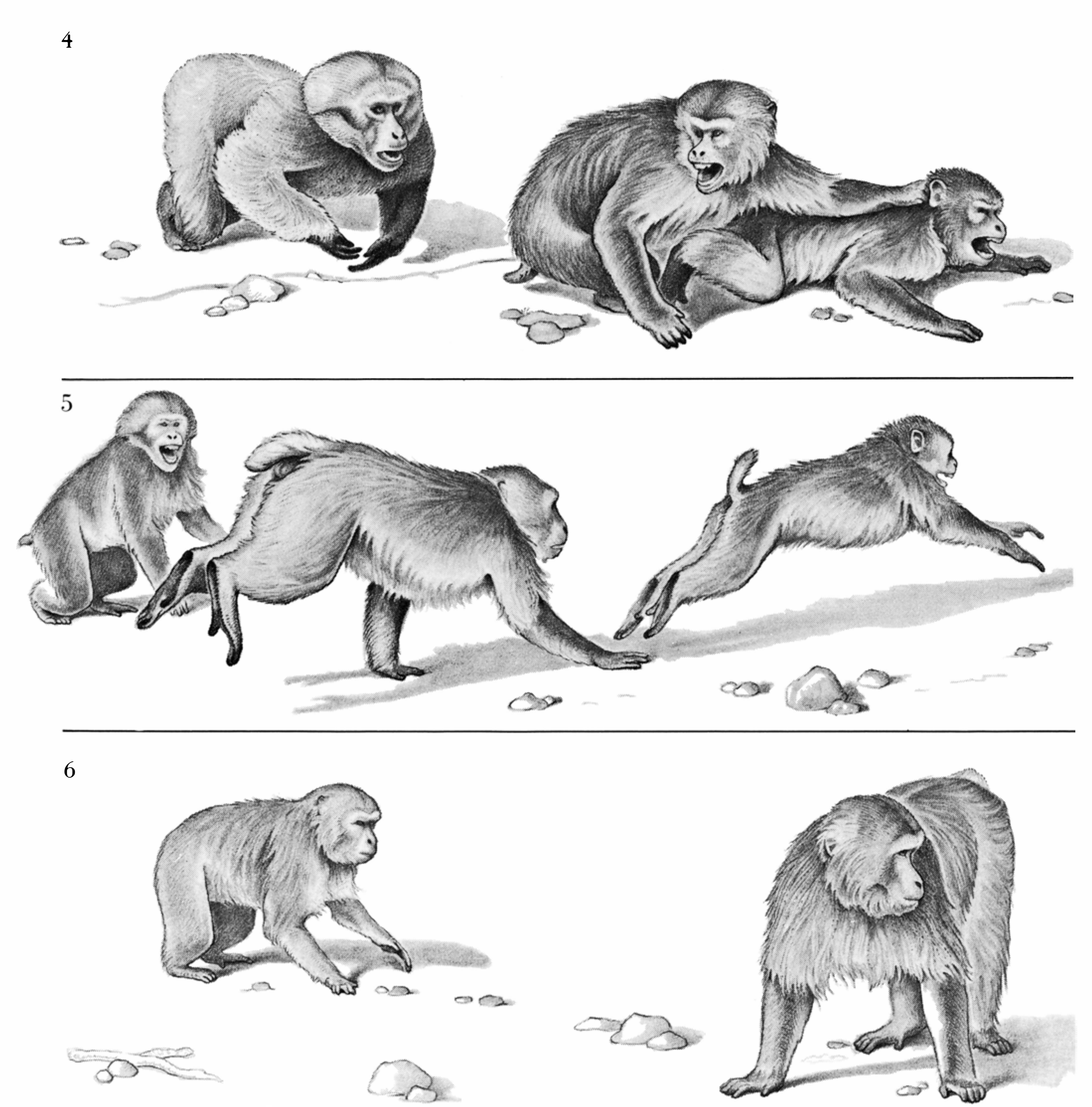


MACHO SUBDOMINANTE CONTROLANDO un incidente de una duración de 1,5 segundos. Los dibujos se basan en las fotografias obtenidas por Kurt B. Modahl con una cámara a motor. La secuencia se inicia cuando la hembra juvenil Tulip amenaza a la hembra adulta Infinity (1). Infinity responde a la amenaza atacando y castigando a Tulip (2 y 3). Los gritos de Tulio atraen a Greater Than, macho subdominante segundo en la jerarquía, quien acude corriendo desde atrás a la izquierda (4) y termina con la pelea persiguiendo a Tulip (5), ame-

centro del grupo; los demás machos son desplazados a la periferia del grupo cuando alcanzan la edad de cinco años, es decir, un año después de llegar a la pubertad. Es probable que, al llegar a esta edad, los machos desplazados deban luchar por conseguir un puesto en la jerarquía periférica, sin contar ya con la ayuda de su madre. Sin embargo, es probable que la influencia de la madre persista, ya que la mayoría de las peleas entre los macacos son ficticias, y es probable que los hijos de madres de rango elevado tengan más confianza en sí mismos que los hijos de hembras de clase baja.

Al sistema jerárquico determinado por la madre se superpone un segundo sistema de organización que, en nuestra opinión, es aún más importante que el anterior para determinar la organización social del grupo: cada animal lleva a cabo una función específica dentro del grupo, que depende de su edad, sexo y rango. Estas funciones pueden ser simples e intercambiables, como el detectar la presencia de depredadores, o complejas y permanentes, como el educar a los hijos, pero, en conjunto, constituyen la base de la organización social, y permiten distinguirla de un simple agrupamiento de individuos.

El papel del macho alfa consiste, al parecer, en dirigir los movimientos del grupo y en defenderlo frente a posibles ataques. Como que la movilidad de nuestro grupo se halla limitada por la presencia de una valla, Arrowhead ha dejado ya de ocuparse de esta actividad,



nazándola luego con la mirada (6). Generalmente, los subdominantes atacan al combatiente de mayor rango; sin embargo, en este caso Greater Than persigue a Tulip, que llevaba la peor parte. Quizá Tulip le amenazó sin proponérselo al iniciar la pelea, el macho se sentía atraído por Infinity, aunque esto sea pura especulación. El comportamiento de los macacos es rico y complejo; no se puede reducir a fórmulas. Su comportamiento varía, según la presencia del macho alfa, las alianzas entre machos y hembras o las simpatías o antipatías individuales.

pero sigue defendiendo al grupo. Cuando, una vez al año, hay que capturar al grupo completo para llevar a cabo la limpieza del corral y reconocer el estado general de los animales, Arrowhead defiende al grupo olvidando su propia seguridad. Dado el pequeño tamaño de la cancela situada en el extremo oeste del corral, nos vemos obligados a separar a los macacos en pequeños grupos, que son capturados sucesivamente, después de conducirlos a lo largo de la pared norte del recinto. Cada vez que uno de estos grupos va a ser capturado, Arrowhead se separa del resto de los animales y se pone a la cabeza del grupo que se dirige a la cancela, enfrentándose a los cuidadores y amenazándoles. Una vez capturado el grupo, Arrowhead se lanza a través de la fila de cuidadores, vuelve con el resto del grupo, y repite de nuevo la acción con cada pequeño grupo de macacos que es capturado.

En los raros casos en que ha sido necesario retirar a un animal herido o enfermo del corral, Arrowhead se ha lanzado a la cabeza de un grupo de machos de escaso rango, junto con uno o dos de los machos subdominantes y algunas de las hembras para atacarnos, mientras el resto de los machos subdominantes permanecía junto a las hembras y a los pequeños situados en el grupo. Los chillidos de los pequeños constituyen el estímulo que con mayor fuerza desencadena las iras de Arrowhead, quien, junto con otros machos adultos, se dirige a los cuidadores gruñendo, balanceando la cabeza y golpeando el suelo con las palmas de las manos.

El papel principal de los machos subdominantes es el de detener las peleas. Generalmente llevan a cabo su misión persiguiendo al más agresivo de los combatientes. Aunque este tipo de respuesta favorece al animal que lleva la peor parte en la pelea, el comportamiento no puede ser calificado de "ético", ya que se produce de forma automática, en respuesta a la dinámica de la pelea. Los chillidos, contorsiones y actitudes del animal derrotado constituyen signos de sumisión que tienden a inhibir la agresividad del macho subdominante, mientras que los gruñidos, boqueos y actitudes de aplastamiento de los pabellones auriculares del vencedor son signos de amenaza que incitan al subdominante a atacar.

En algunos casos, el macho alfa interviene de forma secundaria en el aplacamiento de los combatientes. Cuando el macho dominante se halla presente, la actitud del subdominante es ambigua, alternando las actitudes de amenaza hacia los combatientes con las de sumisión al macho alfa. Si éste ignora la situación, el macho subdominante puede o no atacar al vencedor de la lucha; si amenaza al subdominante, éste se alejará del lugar; en cambio, si coopera con él, ambos atacarán al vencedor del combate.

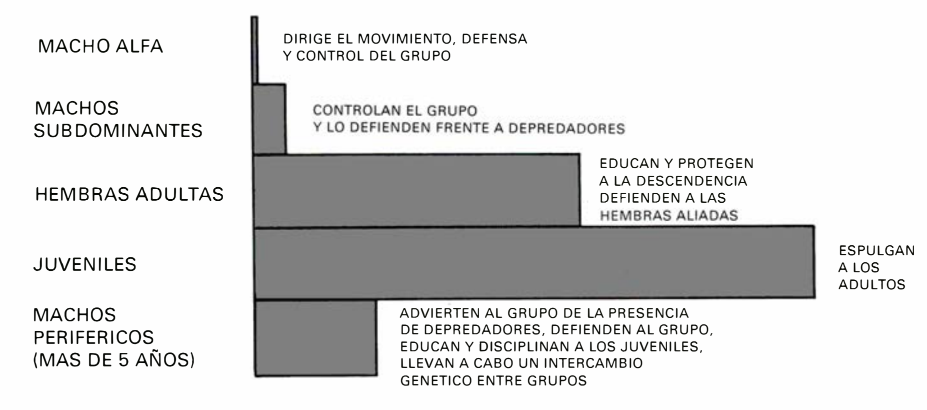
El papel que desempeña cada animal en los grupos de macacos japoneses depende en gran parte de la posición social que ocupa. Sin embargo, cada papel comporta un cierto grado de aprendizaje. Cuando Big Three pasó de ser un macho periférico a ejercer el papel de subdominante al vencer al macho subdominante Bruno, adoptó inmediatamente las funciones de controlar las luchas del grupo, pero su comportamiento presentó diferencias con el de otros subdominantes más experimentados. Cuando se iniciaba una pelea, Big Three se excitaba extraordinariamente, llegando a atacar a animales que no intervenían en la lucha. Sin embargo, al cabo de varios meses aprendió perfectamente su papel, y en la actualidad es uno de los machos más efectivos para controlar los casos de agresividad.

En condiciones naturales, el papel de los machos periféricos consiste, al parecer, en advertir al grupo de la presencia de depredadores, y en ayudar al macho alfa a defender al grupo. También juegan con los machos juveniles, les disciplinan y se dejan espulgar por ellos; a través de este tipo de interacción, los juveniles realizan su aprendizaje sobre el comportamiento social adecuado. Además, los machos periféricos se desplazan de uno a otro grupo de macacos, facilitando así el intercambio genético necesario para que no se produzca una endogamia permanente. En nuestro corral, los machos periféricos no se colocan, en realidad, en la periferia del grupo, ya que el espacio disponible es insuficiente. En general son bien aceptados por el resto del grupo, y rara vez se les molesta.

Las hembras adultas educan y protegen a sus hijos. Los macacos recién nacidos se hallan totalmente indefensos, siendo tan sólo capaces de agarrarse al pecho de su madre; durante los primeros días, la madre sostiene al pequeño con una mano mientras se desplaza. Sin embargo, el proceso de desarrollo es rápido, y al poco tiempo los pequeños intentan ya alejarse de su madre, aunque nunca lo consiguen, ya que la madre posee una gran habilidad para retenerles con sus pies, lo que le deja libres las manos para alimentarse o para espulgar a otro animal. A las dos semanas, la madre empieza a enseñar a su hijo a andar, colocándole en el suelo, apartándose de él y animándole con chasquidos de sus labios. Durante estas fases del desarrollo se ponen de manifiesto ciertas preferencias individuales de las madres. Algunas prefieren llevar a sus hijos montados sobre el lomo. Otras se despreocupan totalmente; nosotros hemos tenido ocasión de observar cómo una madre arrastraba por el suelo a su hijo de uno a dos años, que llevaba colgado de su abdomen.

Los pequeños constituyen un centro de interés para las demás hembras, y en especial para las hermanas del pequeño y para las hembras que no han tenido hijos. Sin embargo, las madres son cuidadosas en extremo con sus hijos, y no permiten que otras hembras agarren al pequeño hasta transcurridas varias semanas del nacimiento. Algunas hembras llegan a recurrir a estratagemas como la de espulgar a la madre hasta que ésta se distrae, aprovechando entonces la ocasión para tocar subrepticiamente al pequeño.

En el momento en que los pequeños empiezan a alejarse de sus madres aparecen las primeras diferencias sexuales en su comportamiento. Los machos juveniles dedican su tiempo a pelear con otros compañeros y, en ocasiones, con algún adulto; las hembras se dedican más al espulgamiento, junto con sus ma-



LA DISTRIBUCION EN CLASES de la organización social de los macacos depende de la edad, sexo y rango de dominancia. Cada una de las clases tiene asignado un papel social bien definido. En la figura se representan las proporciones relativas de cada clase en el año 1974.

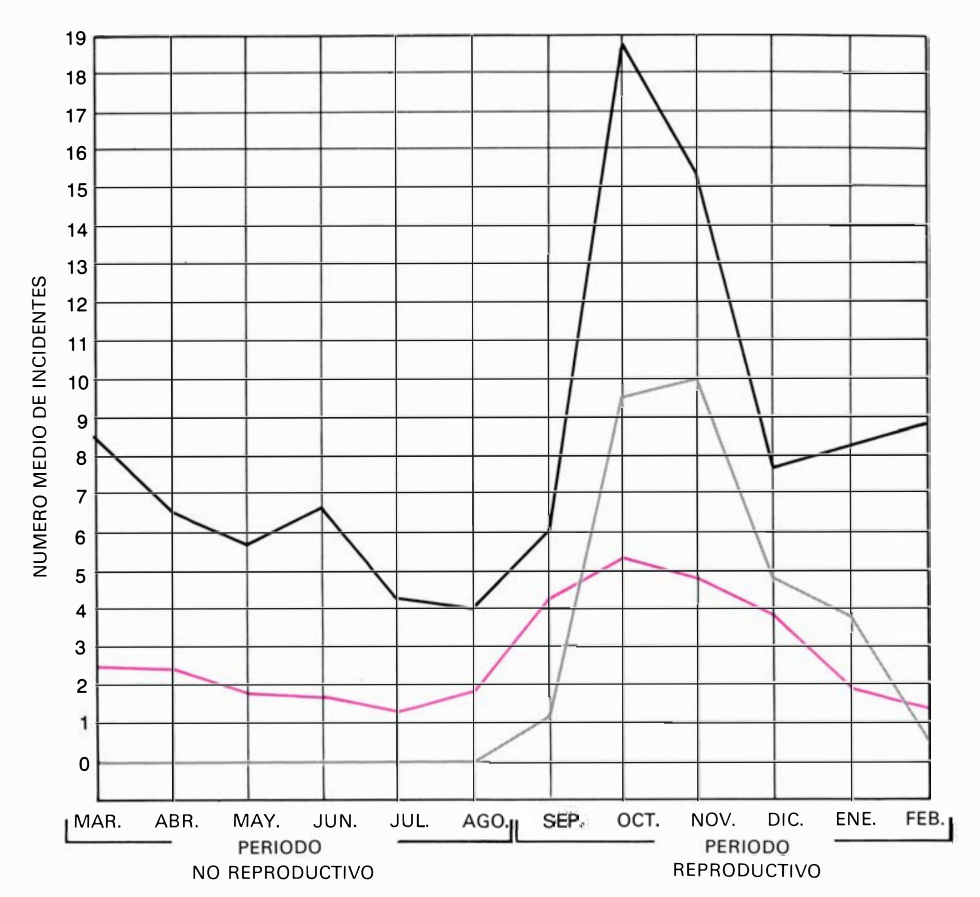
dres y hermanas. En algunos casos, una hembra se une al juego de los machos, aunque generalmente por poco tiempo. Los machos juveniles no se dedican a espulgar a otros animales, pero esta actividad es más frecuente entre ellos que entre las hembras. En caso de que los juveniles espulguen a un animal adulto, éste es casi siempre de su propio sexo.

Los hermanos se defienden entre sí, y también a su madre; sin embargo, después de la pubertad los machos dejan de defender a la madre. Las relaciones existentes entre la madre y sus hijas se mantienen durante toda la vida. Las hembras de una misma familia pasan mucho tiempo juntas, y no es raro ver a dos o tres hermanas adultas espulgando a su madre, mientras sus hijos juegan alrededor.

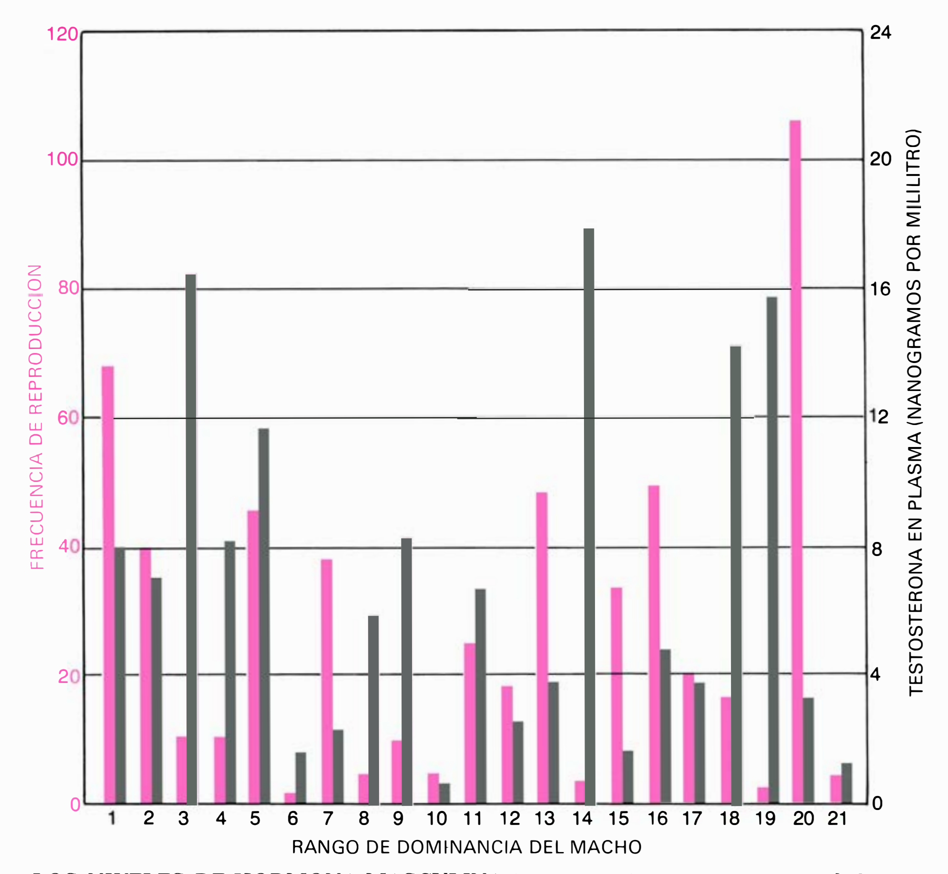
Las hembras adultas forman también alianzas unas con otras, cosa que no ocurre con los machos. Estas alianzas dan como resultado que ciertos machos sean expulsados del grupo en condiciones naturales, o sean relegados a un bajo rango en nuestro grupo. Si un macho ataca a una hembra, otras acuden inmediatamente en su ayuda; en cambio, los machos adultos rara vez se prestan asistencia en situaciones similares. En ocasiones se producen luchas entre grupos de hembras; las peleas pueden ser tan constantes que lleguen a provocar la fisión del grupo, con la partida de varias familias que irán a formar un grupo nuevo. En consecuencia, la organización social de los macacos parece depender más de estas alianzas entre las hembras que del atractivo sexual entre machos y hembras, que en esta especie es estacional y transitorio.

En algunos casos, un macho puede aliarse a una hembra, y defenderla en caso de que sea atacada. En nuestro grupo existe una hembra llamada Gamma que es protegida por dos machos subdominantes. Uno de ellos, Bruno, "adoptó" a Gamma al morir su madre, cuidándola y espulgándola; aún hoy, transcurridos ocho años, Gamma busca protección en Bruno. El hecho de que Bruno no haya copulado nunca con Gamma sugiere que, en los macacos japoneses, el tabú del incesto no tiene origen genético, sino que la prohibición de copular con la madre o con la hermana aparece durante el desarrollo. El otro protector de Gamma, llamado Big Three, sí copula con ella, pero la defiende también durante los meses en que no existe actividad sexual.

La estacionalidad del comportamiento de los macacos es sorprendente. La ac-

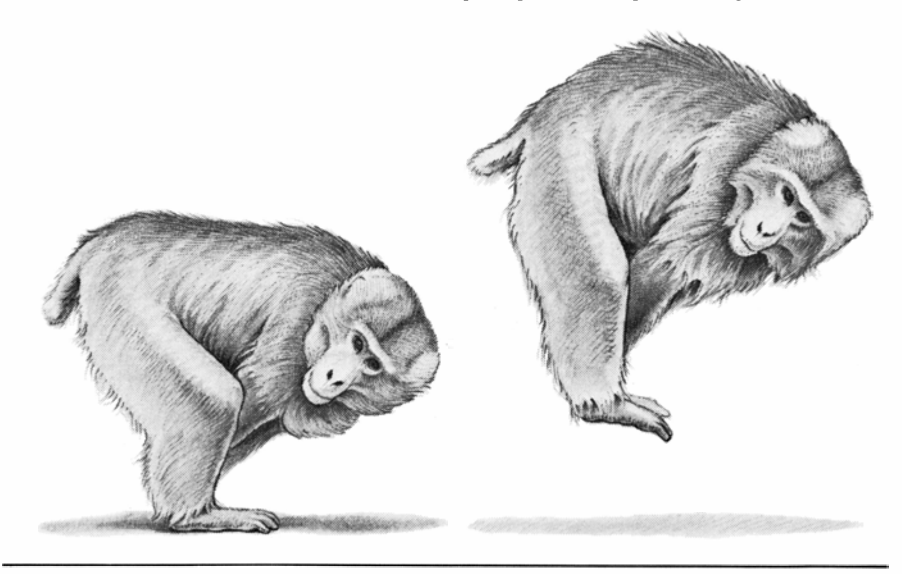


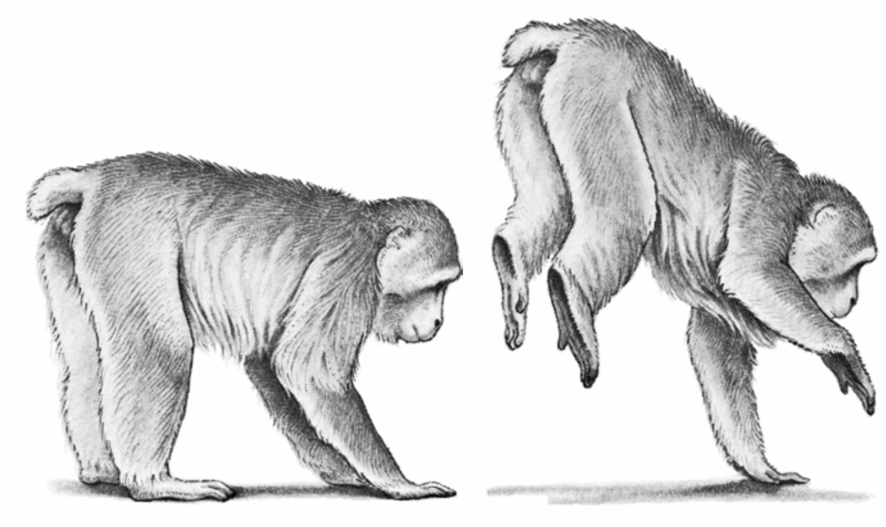
LA DISTRIBUCION ESTACIONAL de ciertas formas de comportamiento es sorprendente: tanto las agresiones por parte de los machos *(color)* como la copulación *(gris)* y las actitudes de cortejo *(negro)* se hallan limitadas, la mayoría de las veces, a las estaciones de otoño e invierno.

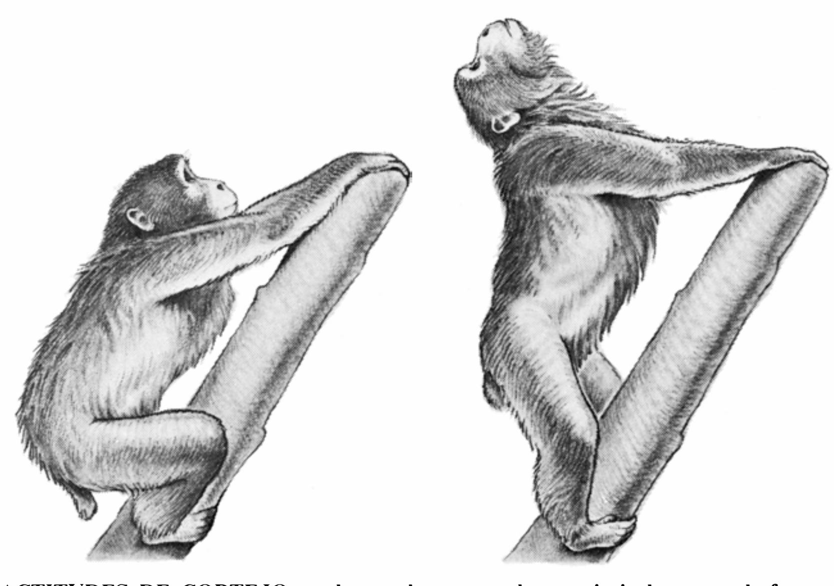


LOS NIVELES DE HORMONA MASCULINA (testosterona) presentes en sangre de los machos *(gris)* no se hallan en correlación con el rango ni con la frecuencia de copulación *(color)*. Aunque es posible que la presencia de niveles elevados de testosterona disminuyan el umbral de agresividad, los factores sociales y de desarrollo poseen mucha mayor importancia con respecto a los factores biológicos en la organización social de los macacos japoneses.

tividad sexual se halla limitada a los meses de otoño e invierno, y es máxima en octubre y noviembre. Los pequeños nacen generalmente en abril y mayo, seis meses más tarde. Durante el verano, las hembras amamantan a sus hijos y se espulgan mutuamente, mientras los machos yacen al sol. Al acercarse el otoño se produce un cambio en el comportamiento, que aparece en primer lugar en los machos. Levantando su corta cola, empiezan a deambular ampulosamente por el corral. Su aspecto posterior llama poderosamente la atención. El brillante pelo blanco situado bajo la cola contrasta con el rojo escarlata de la piel del escroto y del periné. Esta exhibición de colores constituye una actitud de cortejo comparable a la de los pavos reales al desplegar su cola frente a la hembra. Además, los machos sacuden palos y troncos desenraizados, saltan rítmicamente y patean el suelo. Estas actitudes parecen formar parte del rito nupcial, ya que los machos que son más activos a este respecto son los que con mayor frecuencia copulan.



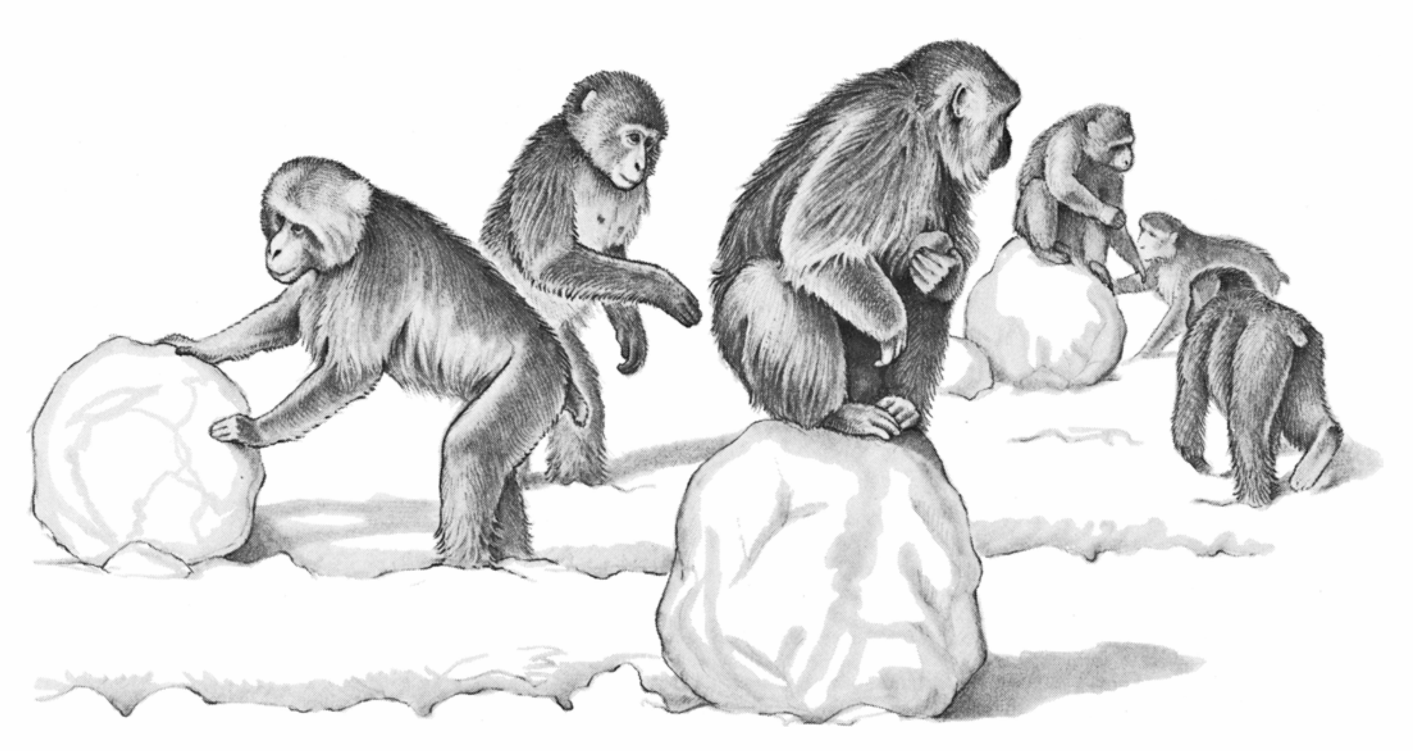




LAS ACTITUDES DE CORTEJO, en los machos, se producen principalmente en la fase reproductiva, correspondientes al otoño y al invierno. Cada macho tiene una actitud preferida, que exhibe en tierra o subido a un tronco. Las más frecuentes son el salto ***(arriba),*** **el coceo** ***(centro)*** **y el corcoveo** ***(abajo).*** **Se ha observado que algunos animales combinan el coceo con el corcoveo.**

Con el tiempo, los machos empiezan a espulgar a las hembras con mayor frecuencia que durante el resto del año, luchan entre sí y también atacan o amenazan a las hembras. Estos ataques no repelen a las hembras, sino que las atraen más y más intensamente, hasta que se forman parejas. Sin embargo, no todos los animales forman parejas, y entre los que lo hacen la relación es totalmente superficial. El macho y la hembra permanecen juntos unas horas, o en algunos raros casos unos días, hasta que la pareja se deshace y cada uno de sus miembros busca un nuevo compañero. Durante la estación reproductiva, cada animal se aparea con varios machos o hembras según el caso.

El rango no tiene relación alguna con el apareamiento, ni en los machos ni en las hembras. Arrowhead, que es el macho alfa, es uno de los más activos, pero Big V, que es uno de los machos de menor rango, ha copulado con mucha mayor frecuencia en los años en que ambos han sido observados conjuntamente. Las hembras de rango elevado copulan también con los machos de menor rango, por lo que el apareamiento parece producirse al azar. Durante el cortejo, el macho es el que adopta el papel activo, pero, al parecer, es la hembra la que elige en último término. Durante estos años hemos tenido ocasión de observar cómo Arrowhead era rechazado por varias hembras mediante el expeditivo proceso de no levantarse cuando él indicaba (mediante un ligero empujón en la espalda de la hembra) que se hallaba dispuesto para copular. En todos estos casos, Arrowhead siguió a la hembra durante varios días, e incluso hasta una semana, antes de decidirse a buscar a una nueva compañera. Mientras, la hembra copuló con uno o más machos de su mejor preferencia. Dado que no disponemos de ninguna prueba para determinar la pa-



LA CONSTRUCCION DE BOLAS DE NIEVE es una manifestación protocultural del grupo de macacos de Oregón. El fenómeno se produjo por vez primera el 13 de enero de 1971, momento en que Big X, macho de escaso rango que había construido una pequeña bola mientras se hallaba comiendo nieve, empezó a hacerla rodar, hasta que la bola alcanzó un diámetro de unos 45 cm. Desde entonces, cada invierno Big X y otros machos construyen grandes bolas de nieve, que constituyen un centro de atracción para el grupo. Aunque esto no tiene relación con la agresión ni con la alimentación (lo que sí ocurre en muchos casos en que usan herramientas) la construcción de bolas de nieve y de otros objetos podría tener un fin adaptativo, preparando a los macacos japoneses para situaciones nuevas y a hábitat diversos.

ternidad, desconocemos qué machos están contribuyendo más intensamente al acervo genético del grupo, pero sí disponemos de datos que indican que, en el momento en que se producen las concepciones, los machos dominantes no intervinieron en el número excesivamente elevado de apareamientos.

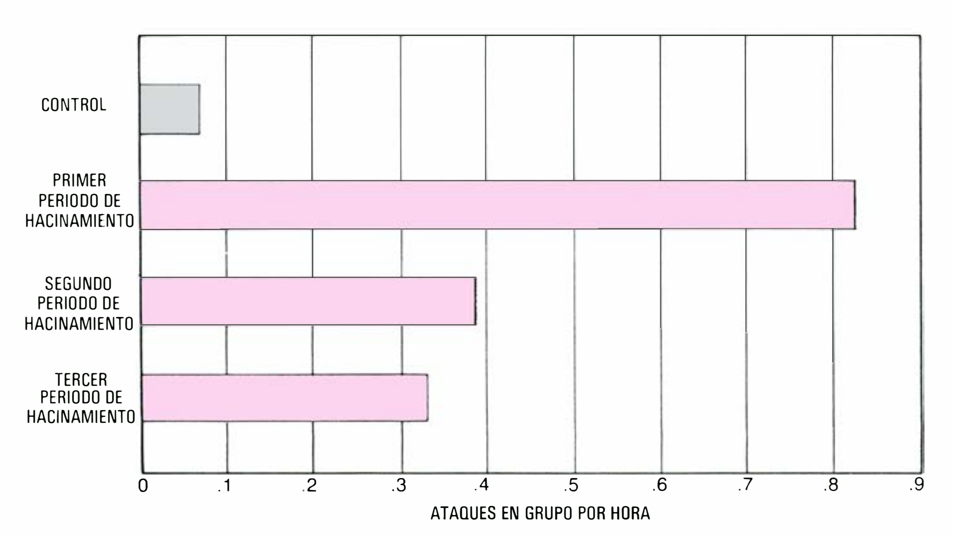
Los cambios que se producen en el comportamiento agresivo de los machos durante la época de apareamiento se deben probablemente a un aumento de la secreción de testosterona. Sin embargo, la relación existente entre la testosterona y la agresividad es compleja, y se halla bajo la influencia de factores sociales. Aunque no se han obtenido determinaciones de los niveles de esta hormona a lo largo del año, hemos llevado a cabo determinaciones de los niveles de testosterona en todos los machos al iniciarse dos épocas de apareamiento consecutivas, comparando los valores obtenidos con el rango del animal. Sorprendentemente, no pudo determinarse la existencia de ningún tipo de relación entre el rango del animal y sus niveles de testosterona. Ello nos ha llevado a la conclusión de que, aunque la testosterona podría disminuir el umbral de agresividad durante la época reproductiva, el comportamiento agresivo se hallaría principalmente bajo la influencia de estímulos sociales.

Al final de la estación reproductiva dejan de producirse apareamientos, y termina también la actividad lúdica de los machos. Los mismos animales que, poco antes, se agredían furiosamente permanecen ahora juntos, jugando y peleando entre sí y con los machos juveniles. Resulta divertido observar a un enorme macho tumbado de espaldas mientras un juvenil le salta encima de la barriga, para, a continuación, bromeando, lanzarse sobre su cuello. Durante este período de tiempo, los machos adultos adoptan una actitud paternal, jugando con los juveniles y espulgándoles, defendiéndoles, cuidándoles e incluso transportándoles sobre sus espaldas. Según Junichiro Itani, de la Universidad de Kyoto, esta actitud paternal, que en nuestro grupo no presente relación alguna con el rango, en el grupo salvaje de Takasakiyama, en Japón, aparece tan sólo en los machos de rango superior.

Estas diferencias no pueden atribuirse tan sólo al confinamiento, ya que es posible que correspondan a diferencias protoculturales. Shunzo Kawamura, de la Universidad de Osaka, que ha estudiado el comportamiento protocultural de varios grupos de macacos en estado natural en Japón, ha observado que cada grupo tiene su propia dieta. Así, por ejemplo, el grupo Shodoshima-O se alimenta normalmente de arroz, cultivado por los campesinos, mientras que el grupo Takagoyama nunca saquea las plantaciones de arroz, a pesar de que las atraviesan con frecuencia durante sus desplazamientos. Kawamura ha visto también a una hembra del grupo de la isla Koshima lavando una papa en un arroyo. Esta costumbre de lavar las papas se extendió, posteriormente, al resto de los animales, hasta que uno de ellos empezó a lavar las papas en el mar, quizá para sazonarlas.

Gordon R. Stephenson, de la Universidad de Wisconsin, ha observado también diferencias en el cortejo entre los grupos Miyajima, Arashiyama y Koshima. Las hembras, durante el cortejo, montan a veces a los machos; sin embargo, los tres grupos presentaron importantes diferencias a este respecto: en el grupo Miyajima sólo las hembras de bajo rango montaron a los machos; en el grupo Arashiyama lo hicieron todas las hembras, y en el grupo Koshima absolutamente ninguna. En nuestro grupo existen algunas hembras que montan a los machos, sin relación alguna con su rango ni con el del macho.

También hemos tenido ocasión de observar una forma curiosa de comportamiento protocultural en nuestro grupo. En invierno de 1970-1971, uno de los machos periféricos llamado Big X construyó una bola de nieve, que arrastró por el suelo hasta que alcanzó un diámetro



LOS ATAQUES EN GRUPO a un mono de bajo rango, en los que intervenían hasta 35 animales, provocándole graves heridas e incluso la muerte, aumentaron en frecuencia cuando el grupo de Oregón fue confinado en una jaula de unos 16 por 16 metros, durante un experimento realizado por Bruce Alexander. Sin embargo, en los tres períodos de hacinamiento que tuvieron que soportar, la agresividad disminuyó progresivamente, lo que indica que los ataques se debían más el traslado de los animales a un ambiente extraño que al aumento de densidad de población.

de unos 45 cm. Desde entonces, cada invierno hay animales que construyen bolas de nieve; estas bolas atraen la atención de los pequeños y de los juveniles, que juegan con ellas, y también para los adultos, que las utilizan para sentarse. Las bolas de nieve no tienen ninguna finalidad, pero, al parecer, han sido el producto de la inteligencia, de la habilidad y de los largos períodos de ocio a que se ven sometidos los macacos.

Como que la cerca que rodea el corral impide que los animales se escapen de un posible agresor, el nivel de agresividad es probablemente superior en nuestro grupo de lo que sería en condiciones naturales. Para solucionar el problema, se han colocado en el recinto pequeños barriles de acero, vacíos, atornillados al suelo, y provistos de una asa a la que puede agarrarse el animal agredido si su agresor intenta sacarle del barril. Estos barriles son utilizados con frecuencia por los animales de menor rango, y han servido también para detener los ataques en grupo.

El por qué, de vez en cuando, un grupo de macacos ataca a un determinado animal sigue siendo un misterio. Los ataques no parecen ser debidos al hacinamiento. Bruce Alexander, que en la actualidad pertenece a la Universidad Simon Fraser, estudió el problema hacinando a los animales de nuestro grupo en una jaula de unos 16 por 16 metros adyacentes al corral, analizando su comportamiento durante tres períodos de hacinamiento y durante tres períodos de control en el corral. Sorprendentemente, el hacinamiento no provocó la destrucción del orden social del grupo; de hecho, la jerarquía de dominancia se mantuvo de forma más estable en la jaula que en el corral. Los actos de agresión y los ataques en grupo aumentaron durante los períodos de hacinamiento, pero su frecuencia disminuyó en los sucesivos períodos en que los animales permanecieron confinados. Ello indica que la mayor frecuencia de los ataques en grupo observada en condiciones de hacinamiento se debería más al traslado de los animales a un ambiente extraño que al aumento de la densidad de población. La hipótesis de Alexander se basa también en la observación de que al trasladar inicialmente a los animales de la jaula en que se encontraba, cuya superficie era de unos 100 metros cuadrados, al corral, que tiene unos 8000 metros cuadrados, la agresividad aumentó también de forma inmediata, manteniéndose en niveles altos tres semanas y provocando la muerte de un animal macho de escaso rango y heridas graves en otros tres.

Nuestros estudios en este grupo se han centrado en el efecto del aumento de densidad de la población tanto sobre la agresividad como sobre la reproducción. En estos momentos seguimos un protocolo parecido al utilizado por John B. Calhoun, del National Institute of Mental Health, quien dejó que su población de roedores se reprodujese en el laboratorio, en un espacio limitado, hasta que aparecieron patrones anormales de comportamiento que abocaron, en todos los casos, a una drástica disminución de la población. También en nuestro caso hemos dejado que la población de macacos aumentase de forma continua dentro de los límites del corral.

Hasta el momento no han aparecido en el grupo muestras de comportamiento anormal, a pesar de su rápido crecimiento, que ha llevado a la población a pasar de los 46 miembros que la constituían en 1964 a los 230 que la componen en 1976. De hecho, el número de enfrentamientos entre machos adultos disminuyó significativamente entre las épocas reproductivas de 1971 y 1972, época en que la población experimentó un crecimiento de un 17 por ciento. (El número de machos adultos en la población siguió siendo de 21.)

Un análisis detallado de la situación permitió comprobar que el aumento de la agresividad experimentado durante la época reproductiva de 1971 se debió a un reajuste de la jerarquía que afectó a los seis machos de mayor rango, mientras que en 1972 la jerarquía se había ya estabilizado. En consecuencia, parece ser que nuestro grupo de macacos se diferencia de los roedores en su respuesta frente al aumento de densidad de la población, probablemente debido a que su estructura social constituye un factor crítico en la regulación del comportamiento agresivo.

Por otra parte, a pesar de que los macacos se hallan confinados en un espacio unas 100 veces menor que su zona de distribución original, que era probablemente de unos 900.000 metros cuadrados, es posible que el nivel de hacinamiento no haya alcanzado aún límites patológicos. Daniel S. Stokols, de la Universidad de Carolina del Norte, ha establecido una distinción entre el proceso físico de densidad de población, definido a partir de parámetros puramente espaciales, y la experiencia del hacinamiento, que constituye una motivación provocada por la interacción entre factores espaciales, sociales e individuales. Teniendo en cuenta esta distinción hemos llevado a cabo observaciones en el grupo de macacos japoneses Arashiyama, que se hallan confinados en un recinto de unos 400.000 metros cuadrados cerca de Laredo, en el estado de Texas. En esta situación hemos comprobado que el grupo, formado por 150 animales, se desplaza en una superficie de 32.000 a 40.000 metros cuadrados al día, cubriendo la superficie total de 400.000 metros cuadrados cada varios días, aunque durante la mayor parte del tiempo, tanto de día como de noche, el grupo se halle situado en una superficie de menos de 8000 metros cuadrados. En consecuencia, es posible que nuestro grupo no se halle aún en situación de hacinamiento, sino simplemente que su distribución se encuentre limitada.

Sin embargo, parece poco probable que la población de nuestro corral pueda seguir creciendo de forma indefinida sin que se produzcan alteraciones patológicas en los animales. En el grupo Gagusan, de Japón, el número de animales se ha mantenido constante, en unos 130 a 160 animales gracias a sus frecuentes escisiones, habiéndose observado que la fisión del grupo va siempre precedida de un deterioro de la estructura social. Es posible que las escisiones se deban a una especial sensibilidad de la estructura social de los macacos frente al aumento de densidad de la población, que habría evolucionado hasta constituir un mecanismo intrínseco de control de la población. Alexander ha señalado la posibilidad de que, una vez superado el umbral de exceso de densidad de la población, se pongan en marcha dentro del grupo interacciones territoriales que den lugar a su escisión, sin que disminuya la fertilidad.

En un futuro próximo esperamos poder poner en marcha un estudio que permita comprobar la hipótesis citada. En estos momentos han terminado las obras de construcción de un corral adyacente e idéntico al anterior, con el que se hallará conectado a través de dos pasadizos, para determinar si, en el momento en que la densidad del grupo ponga en peligro su estabilidad social, se produce una escisión que aumente su zona de distribución. Si la escisión se produce, sin que disminuya la fertilidad del grupo, será posible llegar a la conclusión de que los macacos japoneses poseeen mecanismos de control de la población muy distintos de los que presentan los roedores.

En caso de que se produzca una escisión de nuestro grupo, será también posible estudiar el mecanismo de formación de una nueva organización social. ¿Qué familias son las que van a emigrar? ¿Qué machos adoptarán el papel de subdominantes? ¿Quién será el macho alfa? ¿Cuál será el efecto de la antigua organización social sobre la formación del nuevo grupo? ¿Qué ocurrirá con la antigua organización social si sus miembros más importantes la abandonan?

Además del interés intrínseco que estas preguntas poseen para el estudio de las sociedades de primates, es posible que los resultados que se obtengan de estos estudios permitan comprender mejor la compleja estructura del comportamiento humano. Supervivencia equivale a adaptación, y el conocimiento de los mecanismos de adaptación de los macacos puede contribuir a la supervivencia de la especie humana.

Tempestades de polvo

Aunque en muchas partes de la tierra aparecen raramente, constituyen uno de los rasgos dinámicos más importantes del planeta, pues modifican su superficie e inciden hondamente en los asuntos humanos

Sherwood B. Idso

Un conductor que viaje entre Phoenix y Tucson (Estados Unidos por la autopista interestatal 8 puede toparse con una señal que le avise: "Reduzca velocidad, polvo levantado por el viento". Las señales de precaución, cuyo contenido varía, se repiten cada siete kilómetros, aproximadamente; la autopista cruza un desfiladero donde las tempestades de polvo se dan frecuentemente y a veces reducen tanto la visibilidad que provocan choques. Una tempestad de polvo es siempre un fenómeno meteorológico que impone; y como las señales indican, puede resultar también muy peligroso.

Son varios los procesos que pueden arrojar polvo a la atmósfera. Probablemente, el más dramático de ellos sea el vulcanismo. La erupción de un gran volcán lanza vastas cantidades de detritos que alcanzan hasta bien entrada la estratosfera. A esa altura, los fuertes vientos diseminan rápidamente el polvo sobre la totalidad del globo. Algunos investigadores que han estudiado el fenómeno creen que tales sucesos son la causa de varios períodos de apreciable enfriamiento del clima terrestre.

Las actividades humanas producen mucho del polvo que se arroja a la baja atmósfera. Las áreas urbanas con concentraciones industriales son importantes fuentes de polvo. Pero también en zonas rurales el hombre interviene en ello, debido a los métodos de roturación y quema de maleza usados para limpiar la tierra y a los sistemas de cultivo que dejan el suelo a merced de la erosión por el viento. Cualquiera que sea el origen del polvo, el viento es el factor común que levanta del suelo las partículas de polvo y las reparte a través del aire. En efecto, la definición que da el diccionario de una tempestad de polvo es "vientos fuertes que transportan nubes de polvo".

No es la tierra el único planeta con fuertes vientos y polvo. Cuando el *Mariner 9* se acercaba a Marte a finales de 1971, todo el planeta estaba envuelto en un velo de polvo levantado por una tempestad. Peter J. Gierasch y Richard M. Goody, de la Universidad de Harvard, conjeturaron que tales enormes tempestades de polvo marcianas vienen producidas por un proceso de automantenimiento que convierte energía solar en energía del viento en gran escala y en forma organizada, como en los huracanes terrestres. Marte tiene, asimismo, tempestades de polvo locales.

El estudio de las tempestades de polvo es, probablemente, muy antiguo, pues desde siempre han desempeñado un papel crucial en el bienestar físico y económico del hombre en muchas partes del mundo. En el nivel de interacción más directo, la tempestad de polvo puede amenazar la propia vida; en ocasiones, las personas y los animales encontraron la muerte por asfixia durante una fuerte tempestad de polvo. Por ejemplo, se tienen testimonios de que una importante tempestad de polvo en 1895 causó la pérdida del veinte por ciento del ganado en Colorado oriental.

Se sabe también que las tempestades de polvo diseminan gérmenes patógenos perjudiciales para el hombre, animales y plantas. En el sudoeste de los Estados Unidos, el protozoo parásito *Coccidioides immitis*, que produce la enfermedad humana llamada coccidioidomicosis (vulgarmente llamada "fiebre del valle"), se propaga ampliamente de esta manera.

Los animales, y el ser humano, son atacadas en todo el mundo por el *Criptococcus neoformans*, organismo parecido a la levadura y arrastrado por el viento que ataca el sistema nervioso central, los pulmones y la piel. Por lo que se refiere a las plantas, el viento no solamente arrastra gérmenes patógenos, sino que las hiere de tal modo que permite al germen penetrar en ellas.

La amenaza más directa de las tempestades de polvo contra el hombre es probablemente la erosión del suelo. La tendencia del viento a erosionar el suelo es agravada frecuentemente por las labores de labranza, aunque hay pruebas de distintos tipos que indican que la región de las Grandes Llanuras de los Estados



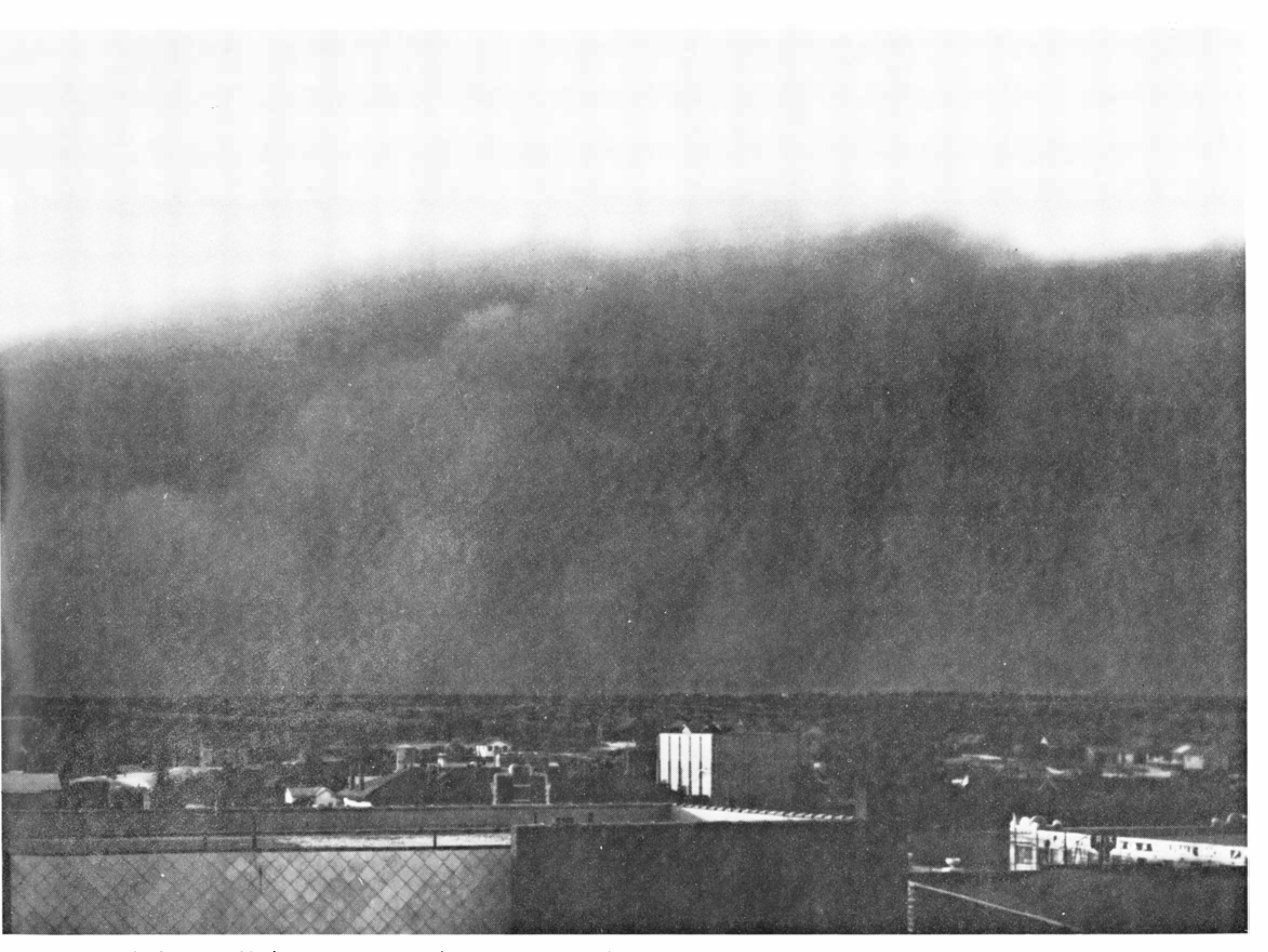
TEMPESTAD DE POLVO abatiéndose sobre la comunidad de Big Spring, Texas, en

Unidos estuvo sometida periódicamente a tremendas tempestades de polvo mucho antes de que se introdujeran los métodos actuales de cultivo y cría de ganado. Sin embargo, la tendencia de la agricultura a incrementar el riesgo ha conducido a los Estados Unidos, y a otros países, a alentar la investigación sobre las características de las tempestades de polvo y el descubrimiento de métodos de defensa. En Big Spring, Texas, donde se han registrado grandes tempestades de polvo, un grupo de expertos del Servicio de Investigación Agrícola del Departamento de Agricultura de los Estados Unidos elaboró un método para predecir, al comienzo de un año, el número de días durante el cual se puede esperar que el viento levante polvo; el margen de error de este método es igual o menor que cinco días en un 72 por ciento de las veces. Este sistema permite a los agricultores tomar medidas preventivas. En la actualidad, el método básico para controlar la erosión por el viento es mantener el suelo cubierto, lo cual suele resultar bastante eficaz. Si la lluvia se mantiene por debajo de lo normal durante dos años y la cantidad de cosecha residual es pequeña, puede recurrirse a la remoción de la tierra como medida de urgencia. La extensión arada se muestra como una superficie rugosa, salpicada de terrones, que tiende a impedir que las partículas queden sueltas porque, éstas, se acumulan en las cavidades producidas entre los terrones.

La llegada de los satélites meteorológicos ha suministrado una perspectiva más ampliada de las tempestades de polvo. Los satélites revelan configuraciones grandes e integradas de movimiento de polvo y también tremendos campos de dunas que permiten explicar el comportamiento y el mecanismo de las tempestades de polvo.

La unidad básica mínima de tempestad de polvo que tiene importancia meteorológica es, probablemente, la causada por una corriente descendente de aire relativamente frío procedente de un cumulonimbo. Cuando esta nube ha crecido hasta el punto en que empieza a producir lluvia, el aire se enfría apreciablemente cuando la lluvia pasa a su través y se evapora total o parcialmente. Al ser este aire más frío que el circundante, también es más denso, de modo que baja en forma de corriente cuya velocidad es aproximadamente proporcional a la altura de la cima de la nube. Cuando este aire frío y denso choca contra el suelo, se desvía y se mueve hacia adelante, precediendo a la parte anterior de la nube en forma de una gran lengua de aire más denso que fluye sobre el suelo. Si éste está seco y cubierto de materia suelta, gran parte de la misma será barrida por la turbulenta y violentamente agitada parte anterior de la lengua de aire.

Una tempestad de polvo de este tipo, que se forma sobre los desiertos durante un período de inestabilidad con-



septiembre de 1930, época en que se llamó al suroeste estadounidense el *tazón de polvo* por sus numerosas tempestades de este tipo. Las llanuras del sur fueron duramente afectadas por tempestades de polvo que desplazaron grandes masas de tierra, degradando así la agricultura.

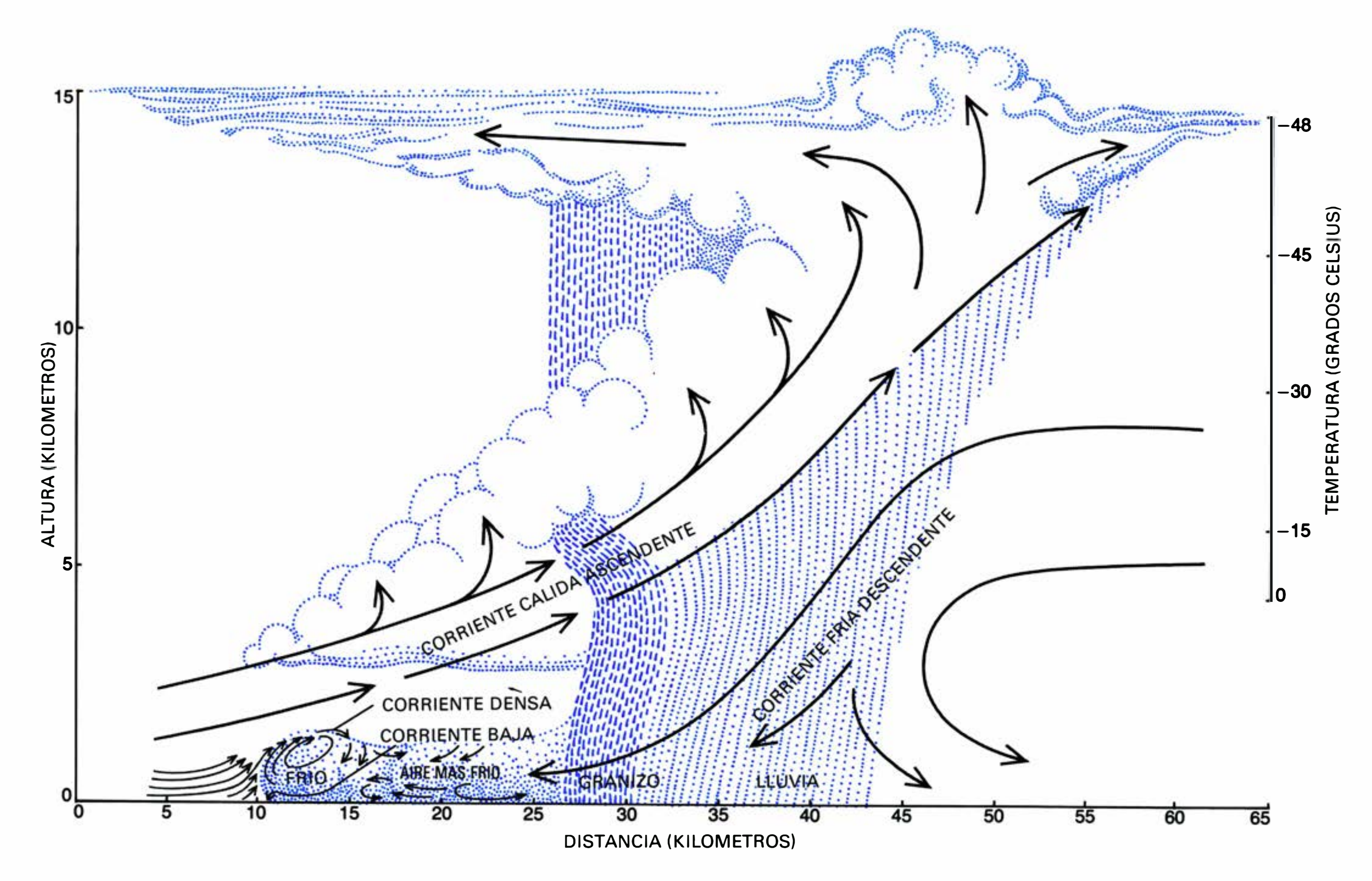
vectiva, se llama un haboob, palabra derivada de otra árabe que significa viento violento. Los haboobs se dan frecuentemente en Sudán, en los alrededores de cuya capital, Khartoum, se registran unas 24 veces al año. Como están asociados a la estación lluviosa en Sudán, se forman generalmente entre mayo y octubre, con un ligero máximo en junio y julio. La cantidad de polvo transportado por los haboobs en Sudán es notable: en dos meses se pueden apilar cuatro o cinco metros de material contra un objeto que esté expuesto a la furia de esas tempestades.

Muy similares son las tempestades de polvo que se dan en el suroeste de los Estados Unidos, en donde se les reconoce por haboobs americanos. En Phoenix, ocurren con una frecuencia de unos 12 al año. El mayor volumen de la investigación sobre tempestades de polvo acometida en los Estados Unidos se realiza en Arizona.

Como en Sudán, la máxima frecuencia de los haboobs en Arizona corresponde a la estación lluviosa (julio y agosto). Se producen con cierta regularidad una vez que este estado ha quedado invadido por el aire marítimo tropical, característico del monzón de verano. Durante la segunda quincena de junio, una corriente húmeda, a gran altura, comienza a invadir Arizona procedente del golfo de México. Durante unas cuantas semanas avanza y retrocede alternativamente, como si se tratara de mareas oceánicas. Sin embargo, a partir de la segunda semana de julio queda bastante fija. Se refuerza, entonces, en los niveles bajos por oleadas periódicas de aire tropical húmedo procedente del Pacífico que asciende a lo largo del golfo de California. Esas oleadas originan frecuentemente largas y curvadas líneas de turbonada, con haboobs extendiéndose en la parte delantera inferior de cada célula en la línea. Las lenguas que se suelen derramar individualmente se sueldan entre sí, formando una suerte de pared sólida de polvo que se extiende a lo largo de cientos de kilómetros.

Las fotografías de satélites han mostrado que los sistemas de tempestades de esta clase se originan con frecuencia al sureste de Tucson, a partir de densas masas nubosas formadas sobre la Sierra Madre Occidental, al norte de Sonora, en México. Las masas nubosas se forman de manera casi explosiva poco después de medio día, comenzando como células aisladas que se mueven lentamente hacia el noroeste. La actividad tormentosa parece originarse en la convergencia del aire húmedo elevado por los vientos que ascienden por las laderas orientales a última hora de la mañana y por las muy recalentadas laderas occidentales durante la tarde. En ese tiempo se desarrollan generalmente nuevas células y forman una línea de turbonada cerca de Tucson. Los sistemas generalmente se intensifican a la vez que avanzan Valle de Santa Cruz arriba, hacia Phoenix. Las nubes en la línea frecuentemente alcanzan los 12.000 metros y no es raro que lleguen a unos 18.000 metros. El polvo levantado por la densa lengua delante de ellas llega a subir hasta la misma base de las nubes, que puede variar entre 2500 y 4500 metros sobre el suelo.

Los haboobs suelen avanzar a una velocidad media de 45 kilómetros por hora. Las rachas máximas en el interior de un haboob, sin embargo, alcanzan unos 85 kilómetros por hora. Con la llegada de un haboob la humedad relativa



EL MECANISMO GENERADOR de una tempestad de polvo del tipo de línea de turbonada se representa aquí esquemáticamente. Las inclemencias atmosféricas, como la lluvia y el granizo, producen una corriente descendente de aire frío que se extiende sobre el suelo y se mueve hacia adelante en forma de corriente densa, es decir, una masa de aire en movimiento que es más pesada que el aire circundante porque es más fría. Esta corriente arrastra consigo polvo y arena de la superficie. Además, el aire frío, denso, impulsa hacia arriba el aire más cálido y ligero que va encontrando, reforzando así la corriente cálida ascendente que crea nuevas nubes de lluvia y renueva el ciclo.

generalmente sube y la temperatura del aire desciende. Se han medido caídas casi instantáneas de temperatura de 15 grados Celsius en unas pocas tempestades fuertes en el momento de llegar la barrera de polvo. No obstante, el descenso medio de temperatura sólo es de unos 7 grados Celsius.

Aunque la visibilidad puede bajar a cero en una tal tempestad, el promedio es de unos 300 metros. Después de la tempestad, ha de transcurrir alrededor de una hora para que la visibilidad vuelva a alcanzar 8 kilómetros; en caso de una tempestad fuerte, la plena visibilidad se retarda casi tres horas. Si la tormenta madre llega detrás de la tempestad de polvo, su lluvia limpia exhaustivamente el aire.

Sin embargo, lo más corriente, es que la tormenta que viene detrás no llegue o su precipitación se evapore antes de alcanzar el suelo, y entonces el polvo quede en suspensión durante horas o incluso días.

El borde anterior de un haboob está siempre cambiando de aspecto. Según J. T. Lawson, de la Universidad de Reading, quien ha estudiado los haboobs en Sudán, "un impulso de aire frío procedente del interior del cuerpo del habood origina la formación de un bulto o lóbulo, que parece crecer hacia adelante a velocidad ligeramente mayor que el promedio del frente. El lóbulo se expande vertical y horizontalmente hasta que, cuando su movimiento hacia adelante relativo al frente ha cedido, empiezan a aparecer irregularidades en su borde anterior que dan lugar a que cualquier nuevo impulso procedente del interior fluya hacia un lado. De esta manera, van apareciendo continuamente nuevos lóbulos a partir de las fases finales de uno anterior".

Con frecuencia he visto ensancharse simultáneamente dos lóbulos adyacentes y encerrar entre ellos un volumen considerable de aire cálido. Los lóbulos fríos en expansión también avanzan encerrando cierto volumen del aire cálido situado delante de ellos. Ambas clases de movimiento pueden crear situaciones inestables, que conducen a una intensa convección. Aunque el torbellino así formado es de vida corta, puede atentar contras las casas que encuentre en su camino.

Tales torbellinos con carácter de tornado se forman no solamente dentro de las tempestades de polvo, sino también delante de ellas, donde el aire cálido es violentamente desplazado sobre el hocico del frente de aire frío que avanza.



LA TEMPESTAD DE POLVO QUE AVANZA presenta en su borde anterior una estructura de lóbulos característica. Este tipo de tempestad se llama haboob, nombre derivado de una palabra árabe cuyo significado es viento violento. La fotografía, aquí representada, está tomada en Phoenix en 1971, y muestra lo que se conoce como el gran haboob de Arizona, Estados Unidos.

A continuación, en la escala de tamaños de tempestades de arena siguen las que se observan desde un satélite. A las fotografías sacadas por los satélites se debe abundante información sobre la extensión de las tempestades de polvo y los lugares donde se originan. En particular, A. A. Grigor'yev y V. B. Lipatov, de la Universidad estatal de Leningrado, localizaron cinco regiones importantes en Africa y Eurasia donde se originan tempestades de polvo.

La primera de esas regiones, situada en el Africa central y occidental, se suele caracterizar por corrientes de polvo de extremada longitud sobre el Sahara meridional. Parecen ser el resultado de una enorme corriente de aire que se mueve en dirección este-nordeste sobre los desiertos arenosos de Mauritania y Nigeria. Bandas de polvo gigantes, de 2500 kilómetros de largo y 600 de ancho cruzan frecuentemente el área asociadas a frentes fríos.

Algunos grandes sistemas han llegado a transportar el polvo, a través del Atlántico, hasta la costa oriental de América del Sur. Las masas de polvo así transportadas habían sido originariamente elevadas cuando los fuertes alisios del nordeste (cuya denominación local es harmattan) se encuentran con el monzón del sudoeste, más frío. Esta convergencia puede ocurrir a una latitud que oscile entre los 20 grados norte en verano y los seis grados norte en invierno. El polvo es elevado a grandes alturas, donde las corrientes que circulan hacia el oeste lo transportan a través del Atlántico.

La segunda región importante donde se originan las tempestades de polvo es la costa meridional del Mediterráneo. Allí las tempestades comienzan con el paso de frentes fríos asociados a vaguadas de baja presión que se prolongan hacia el norte de Africa desde la Europa occidental. En las vaguadas suelen formarse profundas depresiones. Los vientos muy calientes y secos que transportan nubes de arena delante de esas depresiones en movimiento hacia el este se conocen con gran variedad de denominaciones locales.

La tercera región importante se circunscribe al nordeste de Sudán. Grigor'yev y Lipatov describieron, además de la tempestad de polvo tipo haboob que predomina al sur de este área, vastas tempestades de polvo que se forman cuando corrientes frías de aire del noroeste encuentran a los cálidos monzones del suroeste. En tales ocasiones, las masas de polvo pueden elevarse sobre grandes extensiones circundando la región nororiental de Sudán, desde el Nilo hasta el mar Rojo. Los vientos de gran altura transportan frecuentemente este polvo a través del mar Rojo hasta la península Arábiga.

Esta península es la cuarta gran región fuente de tempestades de arena. Se forman éstas de manera algo diferente al mecanismo que las provoca en las otras tres. En la Península Arábiga resultan

del contacto entre dos sistemas que tienen diferentes presiones barométricas; la interacción da lugar a grandes áreas tempestuosas. Tales tempestades se forman generalmente cuando la intensidad de viento aumenta en la periferia occidental de una depresión bárica centrada sobre el sur de Irán. Se forman corrientes cónicas de partículas suspendidas en el aire, que se mueven hacia el sur, y suelen extenderse. En el comienzo de su constitución, que Grigor'yev y Lipatov atribuyen a trombas, hay varias corrientes pequeñas y paralelas, de tres a cinco kilómetros de tamaño. Estas se extienden a lo largo de unos 100 kilómetros, soldándose después en una corriente más poderosa. Las bandas mayores, girando y ensanchándose, se extienden hasta unos 500 kilómetros. Esas corrientes, que se mueven con el alisio, son transportadas por un corredor limitado al norte por las montañas del borde meridional de Asia Menor e Iraq, y al sur por las altiplanicies y montañas de Arabia Saudita.

La última de las cinco regiones considerada fuente de tempestades de polvo, estudiada también por Grigor'yev y Lipatov, circunda el bajo Volga y el Cáucaso septentrional. Las tempestades de polvo en dicha región son también del tipo de zona de tempestad. Se originan en forma típica por aumento del gradiente de presión atmosférica en la periferia noroccidental de lomas de alta presión que se extienden hacia el Volga y Asia central. La causa de tal aumento de gradiente es el movimiento de ciclones que invaden el norte de Europa y la extensión de vaguadas de baja presión al sur de las depresiones.

G. K. T. Ing, de la Universidad de Hawai, describió tempestades de polvo de tamaño similar sobre la China central. Se dan con la máxima frecuencia a principios de primavera, cuando la tierra recién cultivada, la lluvia escasa y los vientos en superficie más fuertes, asociados con frentes fríos, se combinan para crear las condiciones bajo las que el polvo se levanta del suelo. La mayor parte de las tempestades de polvo en China, sin embargo, se producen en los vastos yermos entre 32 y 47 grados Norte y 75 y 115 grados Este, que contiene los grandes desiertos de Takla Makan, Gobi y Ordos y la mayor parte de las tierras de loess en China. Se estima que varios miles de toneladas de suelo son transportadas por el viento cada año en esas regiones, parte de las cuales recorren varios miles de kilómetros con el viento y producen puestas de sol intensamente ro-





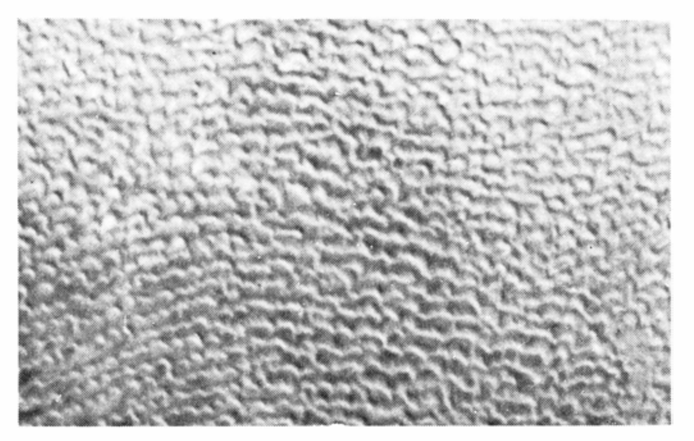
LAS CARACTERISTICAS DE LA SUPERFICIE pueden determinar si se originará o no una tempestad de polvo cuando las condiciones meteorológicas son favorables. Cada una de estas fotografías se tomó en idénticas condiciones meteorológicas: corrientes descendentes de aire fresco creadas por precipitación que se movían por delante de la tormenta madre como corrientes mas densas. Cuando la fotografía de arriba se hizo, en Minnesota, la humedad era alta y había poca materia suelta en la superficie del suelo, por eso no hubo tempestad de polvo. La fotografía inferior fue tomada en Arizona en una época de baja humedad, con suelo seco y pelado; aquí hubo tempestad de polvo.

UNA TEMPESTAD DE POLVO intercontinental se localiza en el Caribe oriental (fotografiada por un satélite meteorológico). La tempestad se originó sobre el Sahara y fue arrastrada hacia el Oeste. Véase la zona gris de punta roma en la parte central derecha de la fotografía. Inmediatamente debajo de la tempestad de polvo, hay un grupo de nubes blancas dispersas que delinean la parte norte de América del Sur.

jas en lugares tan alejados hacia el sur como Hong Kong.

Hay dos tipos diferentes de mecanismos responsables de los principales tipos de tempestades de polvo en China. En el primer mecanismo, el polvo es alzado por los vientos detrás de un frente. Las tempestades que se producen de esta manera suelen ser de tamaño moderado, y el polvo de ellas queda por lo regular confinado a una altura no superior a 1,5 kilómetros. En el segundo mecanismo el polvo se eleva a mucha mayor altura (hasta unos tres kilómetros) por las corrientes verticales delante del frente. El polvo se transporta frecuentemente a distancias del orden de 2500 kilómetros. Las condiciones meteorológicas necesarias





LAS FORMAS DE LAS DUNAS indican las características del viento y por ello ayudan en el estudio de las tempestades de polvo ya que el viento es el factor más importante en un fenómeno semejante. Estas formas de dunas fueron fotografiadas en Arabia Saudita por un satélite de investigación de recursos terrestres. Según un sistema de clasificación elaborado por Edwin D. McKee y Carol Breed, del Servicio de prospección geológica de los Estados Unidos, los tipos de dunas que aparecen en esta página son en rectas paralelas, o lineales *(foto de la izquierda)* **y paralelas onduladas o en creciente** *(derecha).* **Otras formas son radial, en** *U* **y en láminas sin rasgos especiales o en hileras.**

para el primer mecanismo son vientos fuertes y secos soplando en superficie y a alturas moderadas por encima de ella; el segundo mecanismo requiere un vigoroso frente seco con fuerte movimiento ascensional en su parte anterior.

Se puede adquirir indirectamente una cantidad de información sobre las tempestades de polvo estudiando las características de la superficie de los desiertos, por ejemplo, los tipos de dunas. Se conocen algunos de los factores que determinan el tamaño y forma de las dunas: la dirección, fuerza y variabilidad de los vientos, el contenido de humedad del suelo, la vegetación, la topografía subyacente y la cantidad de suelo transportable accesible al viento. El viento es, con mucha diferencia, el factor más importante. El estudio de los grandes mares de arena en los desiertos promete, por tanto, arrojar bastante luz con respecto a las tempestades de polvo en esas regiones.

Como primer paso en este proceso Edwin D. McKee y Carol Breed, del Servicio de Prospección Geológica de los Estados Unidos, han empleado fotografías hechas por satélites para elaborar un sistema de clasificación de dunas que se pueda aplicar a todo el mundo. En él se reconocen cinco configuraciones básicas entre los grupos de dunas: rectas paralelas o configuración lineal, paralelas onduladas o en creciente, estrellada o radial, parabólica o en forma de *U* y sábanas o hileras.

Los grupos de dunas lineales o en rectas paralelas se definen como cuerpos de arena en los que la longitud de las dunas es mucho mayor que su anchura. El área ocupada por las dunas es aproximadamente igual al área entre dunas.

Ambas vertientes de la duna acostumbran ser lo suficientemente abruptas para poder originar avalanchas. Se encuentran formas de dunas de esta clase en el desierto de Simpson en Australia, el desierto de Kalahari en Africa del Sur, la Comarca Vacía de Arabia Saudita y el Sahara.

Los grupos de dunas en líneas paralelas onduladas o en cuarto creciente están constituidas por alineaciones casi paralelas de segmentos en forma de luna creciente. Donde mejor representados están es en el desierto de Kara Kum, en la URSS, y en las Colinas Arenosas de Nebraska, Estados Unidos. Se encuentran variantes de este tipo en el Gran Erg Oriental de Argelia, en algunas partes de Arabia Saudita y en los desiertos de Gobi y Takla Makan, de China.

Las dunas estrelladas o radiales están formadas por diversos brazos que salen de un cono central en distintas direcciones; las dunas se parecen mucho a un gigantesco molinillo de viento. Pueden encontrarse, distribuidas al azar, en partes del Gran Erg Oriental de Argelia y la Comarca Vacía de Arabia Saudita. En otras partes de Argelia y en el Gran Desierto de México se encuentran cadenas de dunas estrelladas.

Las dunas parabólicas o en forma de *U* parecen desarrollarse en regiones donde la vegetación o la humedad, o ambas, tienden a anclar los brazos mientras que el centro es arrastrado por el viento y su arena se mueve hacia adelante. Este tipo de duna se encuentra en el desierto de Thar, en Pakistán, y en White Sands Proving Ground, Nuevo México.

Los términos sábanas o hileras se emplean para designar el tipo de relieve en mares de arena que carece de aspecto geomorfológico definido. La arena se acumula en láminas planas (como cerca de Lima, en Perú) o forma hileras viento abajo que no muestra un relieve apreciable.

El trabajo básico de clasificación de dunas se ha completado ya para los más importantes mares de arena de los desiertos del mundo. La tarea de correlacionar la forma de las dunas con las características del viento y de las tempestades de polvo está aún en curso. El trabajo incluye no solamente campañas en los propios desiertos, sino también investigaciones de laboratorio. Cooperan asimismo las observaciones de las superficies y atmósferas de otros planetas, tales como Marte. Tomados conjuntamente los tres enfoques están comenzando a poner de manifiesto características de estas tempestades de sumo interés.

Aunque se dispone ahora de una buena cantidad de información acerca de las tempestades de polvo, es mucho lo que queda por aprender. El reconocimiento de sistemas en gran escala a partir de fotografías de satélites está aún en su infancia. Cuando se logre, podrá esperarse estar en condiciones de hacer predicciones sobre dónde y cuándo pueden surgir tempestades de polvo. También se necesita hacer más investigaciones sobre características en pequeña escala que son importantes para comprender la erosión del suelo y los torbellinos localizados pero intensos que frecuentemente acompañan el paso de las superficies límites de las lenguas frías y densas. Quedan por hacer descubrimientos de interés, importantes para la ecología vegetal, animal y humana, referentes a la diseminación de agentes biológicamente activos por parte de las tempestades de polvo.

así ven INVESTIGACION Y CIENCIA

«Hace cierto tiempo, pensando en la necesidad de publicar una buena revista de divulgación científica en España, consulté con mi amigo y colega, el bioquímico Alberto Sols, acerca del mejor enfoque a tomar en tal proyecto. Alberto me contestó que lo más práctico sería traducir al español la mejor revista que existe en este campo, es decir, SCIENTIFIC AMERICAN. Afortunadamente, gracias a la Editorial Labor, esto ya es una realidad. Y también un gran acierto, porque la supervivencia del hombre en la tierra en los próximos siglos dependerá en gran parte de que éste reciba una buena formación científica y humanística.»

JUAN ORO

Profesor de la Universidad de Houston, director científico del Proyecto Vikingo de la NASA.





«Soy lector, desde hace muchos años, de SCIENTIFIC AMERICAN, que es modelo en el difícil arte de la divulgación científica de altura. Por ello, considero que INVESTIGACION Y CIENCIA contribuirá a crear el clima social que necesita el desarrollo de la investigación entre nosotros, dando a conocer además la gran tarea ya realizada por los científicos españoles e hispanoamericanos.»

EDUARDO PRIMO YUFERA

Presidente del Consejo Superior de Investigaciones Científicas.

«Me complace enormemente la noticia de la publicación de INVESTIGACION Y CIENCIA, edición en español de SCIENTIFIC AMERICAN, la revista de mayor prestigio en el mundo de la ciencia, sin otro norte que el avance de ésta y su incidencia en la vida del hombre. Qué duda cabe que ha de ser un vigoroso instrumento de progreso para los países de habla española.»

JOSEP TRUETA

Profesor Emérito de Cirugía de la Universidad de Oxford.





«Pienso que acaso la más radical de las deficiencias históricas del alma española sea la escasez de su formación científica, la carencia o la debilidad de los hábitos mentales en cuya virtud es posible hacer ciencia propia y estimar adecuadamente la que los demás han hecho. Mídase, pues, según esto, la importancia que atribuyo a la publicación de la revista INVESTIGACION Y CIENCIA. En ella veo un eficaz recurso para acercarnos al siglo XXI a la altura que éste va exigir a todos los hombres.»

PEDRO LAIN ENTRALGO

Catedrático de la Universidad de Madrid. Director del Instituto Arnau de Vilanova del Consejo Superior de Investigaciones Científicas.



«Tengo muchas razones para creer que la edición en español de SCIENTIFIC AMERICAN va a representar un acontecimiento de real importancia en la vida intelectual española. Entre otras, el uso continuo que de SCIENTIFIC AMERICAN hacemos la mayoría de los que, sabiendo inglés, trabajamos en el campo de las ciencias humanas. La calidad y claridad de los trabajos, así como su amplia y actual temática, hacen de esa publicación una herramienta de trabajo en verdad indispensable, que ahora estará al alcance de todo el mundo.»

JOSE LUIS PINILLOS

Catedrático de Psicología de la Universidad Complutense, Madrid.

La química percloroorgánica

Como sucede con la química de los hidrocarburos, los cloruros de carbono están resultando ser los precursores de una química orgánica imagen de la convencional no sólo insólita, sino incluso sorprendente

Manuel Ballester

La filosofía de elegir como línea de investigación la apertura de un campo de la química esencialmente imaginario y, por ello, inexplorado, con un número ilimitado de compuestos y materiales en potencia, ha sido y sigue siendo uno de los objetivos de mayor alcance de la investigación científica. Si la elección del campo está de acuerdo con una realidad escondida, el momento se halla metodológicamente maduro, se cuenta con los adecuados medios humanos y materiales y se aplica la debida habilidad y diligencia, la probabilidad de obtener en su desarrollo resultados de rango superior, de descubrir fenónos inesperados e insólitos, e incluso auténticamente fundamentales, es óptima.

Si bien este concepto está reconocido en las esferas científicas más competentes como aquel que ha producido y puede causar espectaculares saltos en el discurrir del progreso científico y tecnológico, no obstante, debido a la creencia generalizada, aunque falsa, de que tal enfoque es privativo de superinvestigadores o genios, debido también a la impaciencia de los investigadores relativamente noveles por querer dar fe de su existencia en el mundo científico, a los escepticismos y a las incomprensiones ante la perspectiva de una novedad intangible por parte de promotores y patrocinadores, a la presión de una sociedad de consumo y de sus representantes por lograr resultados poco menos que inmediatos y conducentes a la introducción de nuevos productos y a la mejora de los existentes y, particularmente, al riesgo profesional que entraña y el temor a la soledad científica, dejan actualmente poco margen para la adopción de tal filosofía.

El presente artículo se refiere precisamente a uno de tales campos, cuya apertura efectiva se debe fundamentalmente a los esfuerzos de unos investigadores del Instituto de Química Orgánica Aplicada de Cataluña (C. S. I. C.) que aún no han tenido tiempo para ventilar adecuada y ampliamente en la literatura científica sus hallazgos, subyugados por sus propias exploraciones; científicos que han tenido tal vez la osadía de infringir la norma universal de "publish or perish" (publica o muere) tan venerada en los últimos tiempos.

Las substancias orgánicas, entre las que se encuentran los constituyentes de la materia viva y que determinan esencialmente sus procesos; son, como es sabido, compuestos que derivan de los *hidrocarburos* a través de un proceso real o imaginario de *substitución*. Los hidrocarburos son, a su vez, compuestos que están constituidos exclusivamente por los elementos químicos carbono e hidrógeno. La molécula orgánica en general está formada por un esqueleto de átomos de carbono revestido principalmente por átomos de hidrógeno unidos a aquéllos por fuerzas de enlace químico.

La capacidad poco menos que única de los átomos del carbono de concatenarse entre sí ilimitadamente formando tales esqueletos se debe a sus características electrónicas especiales, en particular a su grado de capacidad de enlace químico; es decir, su *tetravalencia*. A ella se debe, pues, el fenómeno vital tal como lo conocemos en nuestro planeta y, por añadidura, gran parte del desarrollo tecnológico del último centenario (advenimiento de los materiales plásticos. fármacos. insecticidas, etc.).

El hidrocarburo más sencillo, el metano, de fórmula CH_4, posee un esqueleto formado por un sólo átomo de carbono, con el que se enlazan cuatro átomos *monovolantes* de hidrógeno.

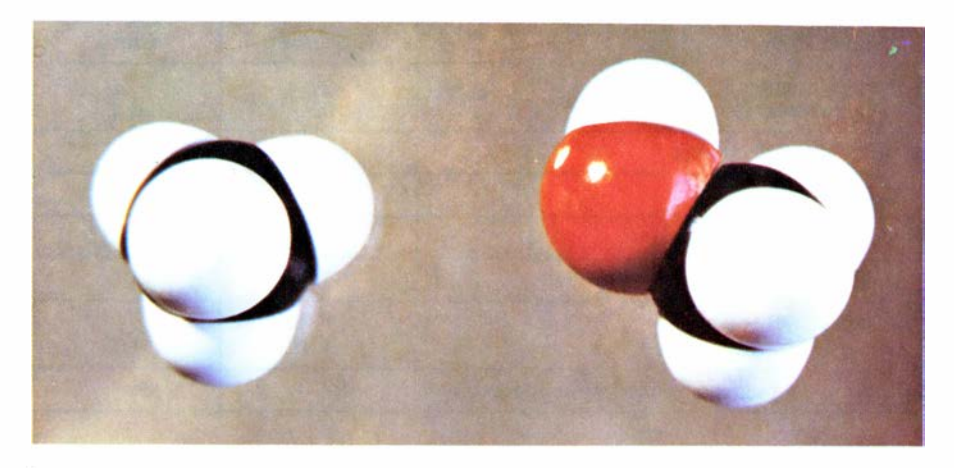
Algunos elementos químicos exhiben la misma capacidad de enlace que el hidrógeno; es decir, son *monovalentes*. Entre ellos, los elementos de la familia de los halógenos: fluor- cloro, bromo e iodo. Por ello, de no existir otras limitaciones, los átomos de estos elementos podrían en principio substituir parcial e incluso totalmente (persustitución) a los de hidrógeno en un compuesto orgánico. En este último caso, podría resultar toda una química orgánica paralela o imagen de la química orgánica convencional que, análogamente a ésta, podría considerarse derivada de los *halocarburos* (o haluros de carbono). Y, efectivamente, se han logrado impresionantes avances en la sustitución total por el elemento flúor que han conducido, por ejemplo, a materiales de tan notables propiedades como el teflón y los freones.

Sin embargo, todos los múltiples intentos para desarrollar substancialmente otras químicas perhaloorgánicas han fracasado hasta hace relativamente poco. En el caso del cloro, si bien en 1820 se registró un comienzo realmente impetuoso y prometedor con la síntesis casi simultánea de nada menos que tres cloruros de carbono: el hexacloroetano (C_2Cl_6), el tetracloroetileno (C_2Cl_4) y el hexaclorobenceno (C_6Cl_6), los dos primeros por Faraday y el último por Julin, que corresponden a los hidrocarburos etano, etileno y benceno, no obstante, esta química ha permanecido prácticamente dormida durante cerca de siglo y medio. No han faltado argumentos teóricos en razón de los cuales la química percloroorgánica sería un campo desierto y yermo (G. W. Wheland, 1960; J. C. Tatlow, 1963), ni sátiras que ridiculizaban el proceso de percloración (sustitución total por cloro) (F. Wöhler, 1840).

Deben mencionarse, sin embargo, los valiosísimos esfuerzos realizados principalmente por dos escuelas, una en Estados Unidos (E. T. McBee) y otra en Alemania (A. Roedig). La primera consiguió ciertos resultados de interés práctico,

mientras que la segunda se ha concentrado en el difícil estudio de compuestos alifáticos perclorados. Los resultados logrados por ellas, tanto los positivos como los negativos, han allanado ciertamente el camino de las investigaciones que se refieren a continuación.

Es muy significativo que la fortaleza de la unión (energía de enlace) entre dos átomos de carbono en una cadena (83 kcal/mol) sea comparable con las 81 kcal/mol de un enlace carbono-cloro (C-Cl). Es sabido que tales cadenas son muy resistentes a quebrarse en procesos químicos; es decir, que en la conversión de un compuesto orgánico en otro aquéllas no suelen alterarse. Por consiguiente, las uniones C-Cl poseen, en ciertos casos, una gran resistencia o estabilidad frente a agentes químicos.



MODELOS A ESCALA de las moléculas del metano y del metanol. El metano es el hidrocarburo más sencillo, de fórmula CH_4; es decir, posee un esqueleto formado por un solo átomo de carbono *(negro)* y cuatro átomos de hidrógeno *(casquetes blancos)*. El metanol *(derecha)* es un derivado de substitución de un átomo de hidrógeno del metano por un substituyente hidróxilo, de fórmula CH_3OH; en el lugar substituido tiene, pues, un átomo de oxígeno *(en rojo)* y otro de hidrógeno *(en blanco)*. La unión con el carbono se produce a través del átomo de oxígeno.

Una importantísima diferencia entre el hidrógeno y el cloro reside en el gran volumen de los átomos de éste, unas cinco veces mayor que el volumén atómico de aquél, lo que ha constituido el obstáculo primordial en el desarrollo de la química percloroorgánica. El lector puede hacerse una imagen de esa dificultad si intenta introducir cuatro pelotas de tenis en un bote de tres. Otras dificultades se deben a que los procesos químicos de substitución de hidrógeno por cloro suelen ser de naturaleza *electrófila;* es decir, que ocurren con agentes químicos que tienen una afinidad intrínseca por los electrones, y por tanto desempeña un papel importante la densidad electrónica sobre los átomos de carbono en que se realiza la substitución. El cloro disminuye tal densidad, por lo que, en líneas generales, el proceso de substitución es tanto más difícil cuanto mayor es el número de átomos de cloro presentes.

Hace unos cuantos años, el autor, fascinado por el fenómeno del impedimento estéreo (impedimento que ocurre en los procesos y las reacciones químicas y que se debe a la insuficiencia de espacio), se propuso demostrar con los recursos entonces disponibles y usando los agentes de introducción de cloro más potentes, que el perclorotolueno (u octaclorotolueno) –el cloruro de carbono alquilaromático más sencillo–, de fórmula C_7Cl_8, resultaba imposible de sintetizar, por cuanto no existe suficiente espacio en la molécula para alojar ocho átomos de cloro, pero sí se podían introducir siete. En efecto, las cloraciones exhaustivas del tolueno daban sea el αH-heptaclorotolueno o bien el oH-heptaclorotolueno, cuya fórmula respectiva era C_7HCl_7. Forzando mucho las condiciones de reacción para lograr la substitución del último hidrógeno por cloro, se producía la *clorolisis;* es decir, fragmentación de la cadena de carbono dando hexaclorobenceno y tetracloruro de carbono.

La realización experimental del proyecto se encomendó a nuestro antiguo colaborador C. Molinet. Por una distracción en la realización de una reacción crucial ésta evolucionó incontroladamente dando una masa negruzca, que normalmente se hubiera echado al vertedero. No obstante, en aquella ocasión, más bien por razones relacionadas con nuestro estado anímico que por consideraciones científicas, se decidió trabajar a fondo aquella masa, obteniendo así un rendimiento substancial de perclorotolueno, la substancia a la que no dábamos probabilidad alguna de existencia y que había eludido multitud de intentos desde que Beilstein se interesó por ella en 1869.

La síntesis del perclorotolueno entrañó el descubrimiento del reactivo de cloración aromática más potente pero a la vez suave (selectivo) que se conoce, denominado *reactivo BMC,* sigla que corresponde a las iniciales de sus autores (Ballester, Molinet, Castañer). En su composición *inicial* entra cloruro de sulfurilo, monocloruro de azufre y cloruro alumínico.

¿Cómo fue posible efectuar la síntesis del perclorotolueno si los modelos a escala indican la improbabilidad de su existencia? La explicación estriba en que tales modelos dan estructuras geométricas rígidas, mientras que las moléculas poseen cierta capacidad de deformarse bajo la influencia de las repulsiones entre sus átomos. Por consiguiente, si bien en la molécula del perclorotolueno existe una gran respulsión entre dos átomos de cloro de la posición *alfa* (unidos directamente al carbono metílico) y los dos de las posiciones *orto* (posiciones colindantes), un conjunto de flexiones de los enlaces C-Cl y C-C permiten sobrellevarla. Así y todo, la energía *de deformación o distorsión es elevada,* probablemente superior a las 20 kcal/mol, valor insuficiente para la rotura de un enlace pero suficientemente elevado para facilitarla.

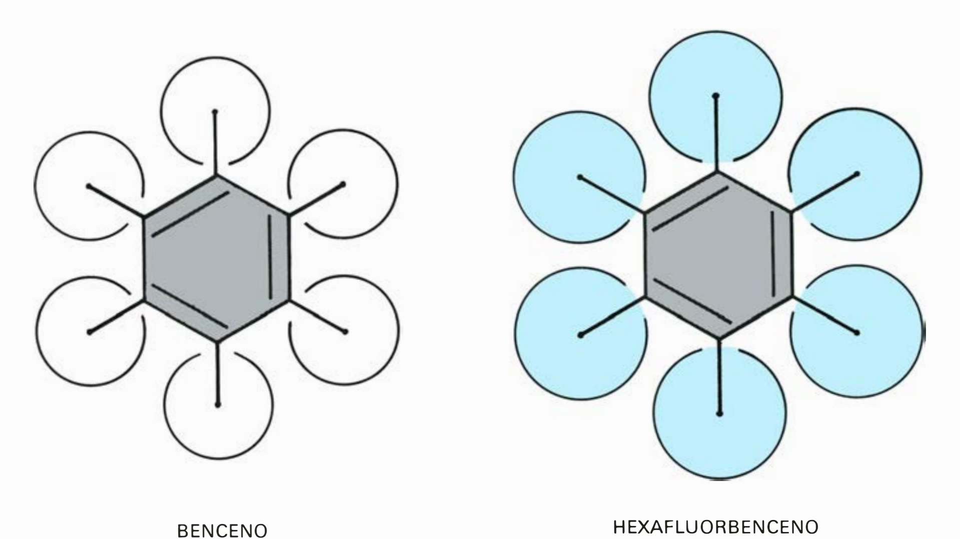
La existencia de una energía potencial de deformación, análoga a la de un resorte en tensión, explica el porqué el perclorotolueno y otros cloruros de carbono alquilaromáticos sean relativamente inestables química y térmicamente, tendiendo a convertirse en compuestos carentes de tensión. En efecto, el perclorotolueno empieza a descomponerse, fragmentándose su molécula, a la temperatura de 185°, mientras que es habitual que las substancias de alto contenido en cloro resistan temperaturas de 300° C, e incluso superiores. Explica, además, las grandes dificultades que presentaba su obtención, puesto que bajo las condiciones drásticas empleadas sufría clorolisis.

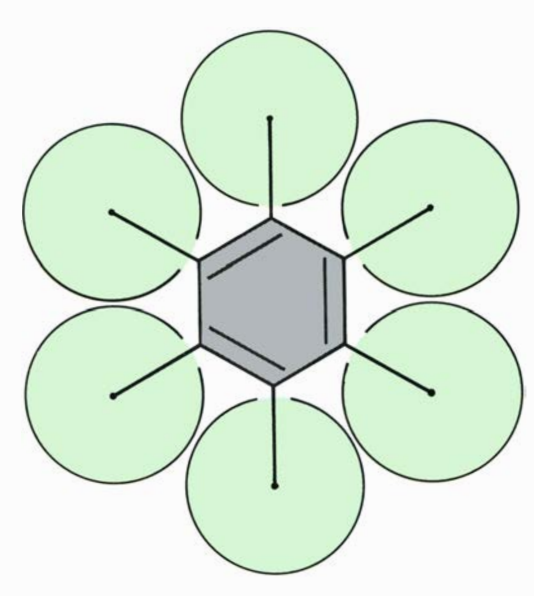
La distorsión molecular origina fenómenos espectrales anómalos. Por iluminación, parte de las moléculas constituyentes de la materia, que normalmente se encuentran en el *estado fundamental,* con una energía electrónica E_o, absorben radiaciones de ciertas frecuencias ν pasando a otros estados, denominados *estados excitados,* de energías E_e. Es condición necesaria para que se produzca tal absorción que se cumpla la conocida condición de Bohr

$$\nu = \frac{E_e - E_o}{h}$$

Propiedades	Hidrógeno	Cloro
Abundancia en la superficie terrestre (%)	0,9	0,2
Número de electrones	1	17
Valencia	1	1
Masa atómica	1	35 y 37
Volumen atómico (Å³)	2,7	14
Longitud del enlace con el Carbono (Å)	1,07	1,77
Energía del enlace con el Carbono (Kcal/mol)	99	81

TABLA COMPARATIVA de propiedades de los elementos hidrógeno y cloro. Una importantísima diferencia entre el hidrógeno y el cloro reside en el gran volumen de los átomos del cloro, que viene siendo unas cinco veces mayor que el volumen atómico de aquél. Tal extraordinario volumen ha sido el obstáculo principal en el desarrollo de la química percloroorgánica.





REPRESENTACION BIDIMENSIONAL a escala de las moléculas del benceno (C_6H_6), del hexafluorbenceno (C_6F_6) y del hexaclorobenceno (C_6Cl_6). Dado el gran volumen del cloro, los esferoides que lo representan casi llegan a tocarse. No se pueden construir modelos a escala del hexabromobenceno o del hexaiodobenceno, pues el bromo y el iodo poseen un volumen atómico mayor.

La mayoría de los compuestos orgánicos son incoloros; es decir, no absorben en la zona visible del espectro de las radiaciones electromagnéticas sino en la zona ultravioleta.

Como quiera que el estado fundamental es menos flexible (mayor carácter enlazante) que el estado excitado, la energía de una transición determinada –como por ejemplo entre los estados A_{1g} y $^1B_{2u}$ de los derivados bencénicos– es substancialmente menor en las moléculas distorsionadas, lo que causa un desplazamiento anómalo de la absorción hacia frecuencias menores (corrimiento batocromático) *(véase la ilustración de la pág. 95)*. El desplazamiento causado por la distorsión molecular en el percloro-*p*-xileno, es realmente espectacular. Así como el hidrocarburo correspondiente, el *p*-xileno, absorbe a una longitud de onda de unos 274 nm (frecuencia de 36496 cm^{-1}), la absorción del percloro-*p*-xileno se centra a unos 365 nm, invadiendo parte de ella la zona visible del espectro. Por ello, este compuesto es amarillo, cualidad excepcional, por cuanto el color suele estar asociado al *fenómeno de resonancia* o deslocalización electrónica.

Mediante el reactivo BMC nos ha sido posible obtener multitud de cloruros de carbono desconocidos, tanto aromáticos como alquilaromáticos, los cuales han rendido a su vez centenares de derivados, incluyendo nuevos cloruros de carbono y otros representantes de diversos capítulos de la química orgánica, tales como percloroéteres, percloroácidos, percloroesteres, perclorocetonas, percloroquinonas, percloroaminas, percloroamidas, etc.

Frecuentemente, los compuestos perclorados poseen propiedades insólitas, e incluso sorprendentes, habiéndose logrado en el curso de su estudio hallazgos de carácter fundamental que trascienden a la química orgánica convencional. Algunos datos que hablan por sí mismos: mientras que el metano (CH_4), por ejemplo, contiene en peso un 75 por ciento de carbono y un 25 por ciento de hidrógeno, el tetracloruro de carbono (CCl_4) está constituído sólo por un 8 por ciento de carbono y el resto de cloro. Simplemente por esto sería razonable esperar que la química percloroorgánica condujera a resultados diferentes.

Muchos perclorocompuestos tensos poseen por lo menos un carbono de configuración tetrahédrica, es decir, dotado de la capacidad de unirse a *cuatro* substituyentes voluminosos. Por ejemplo, el perclorotolueno y el percloro-*p*-xileno poseen uno y dos carbonos tetraédricos, respectivamente.

Otras configuraciones del carbono en la molécula orgánica (hibridaciones) son el *carbono trigonal* y *el carbono digonal* que sólo pueden unirse a tres y dos sustituyentes, respectivamente. Por consiguiente, su capacidad de dar moléculas tensas, debido al apiñamiento de substituyentes, es mucho menor. No obstante, existen compuestos percloroorgánicos con carbonos de hibridación trigonal exclusivamente que presentan distorsión molecular, como el perclorofenantreno y otros derivados de sistemas aromáticos policíclicos condensados.

Sus espectros de absorción electrónica (espectros ultravioleta y visible) son también anómalos por la razón indicada anteriormente.

Hasta este punto se han expuesto y razonado las dificultades principales que impedían una auténtica apertura de la química percloroorgánica antes del descubrimiento del reactivo BMC, así como su superación. ¿Qué encierra además esta nueva química?

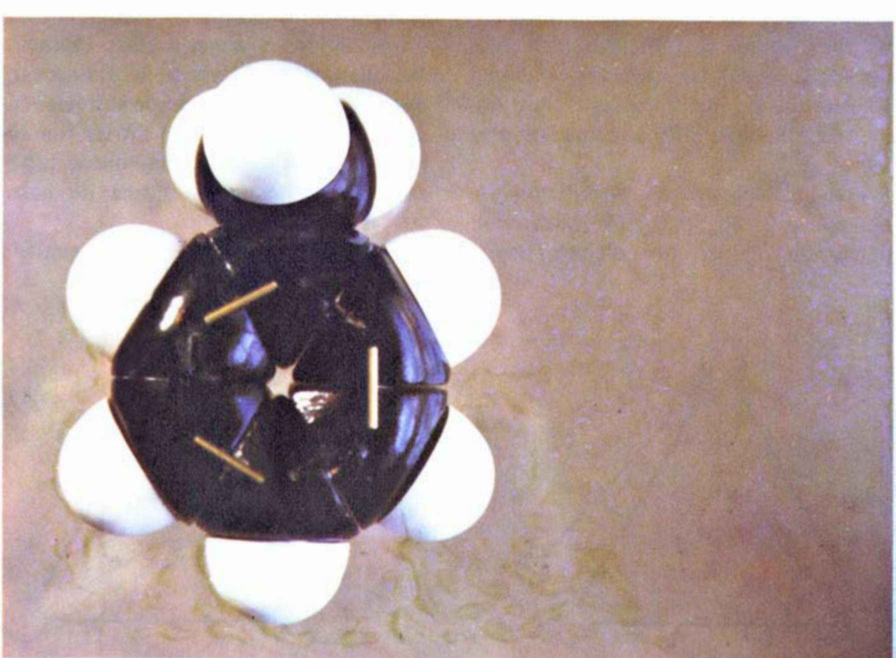
Aparte de los efectos electrónicos debidos a la acumulación de átomos de cloro, que llegan a ser muy importantes, una de las causas que hacen de ella una química diferente es el fenómeno de protección estérea total de la molécula orgánica, puesto que la envoltura de átomos de cloro actúa como una auténtica armadura molecular. Tal armadura reúne los tres requerimientos siguientes que la hacen especialmente eficaz: 1) Es resistente, por cuanto los átomos de cloro unidos a carbonos de hibridación trigonal (carbonos etilénicos y aromáticos) son pasivos frente al ataque químico directo. 2) Es extensa, puesto que el cloro es voluminoso y sustituye totalmente a los hidrógenos, pudiendo cubrir perfectamente las partes vulnerables de la molécula (centros de reacción). 3) Está sólidamente unida al esqueleto de carbono de la molécula, ya que la unión entre el carbono trigonal y el cloro es fuerte.

Estas cualidades no las reúne ningún otro elemento de la familia de los halógenos. En efecto, si bien el flúor cumple las condiciones 1) y 3), sus átomos son demasiado pequeños para brindar una protección eficaz. Los átomos de bromo y de iodo, aparte de que su volumen no permite la persustitución por ellos, no reúnen aquellas dos condiciones.

Dos ejemplos sencillos de estas moléculas blindadas o "moléculas pánzer", como se les ha llamado, son el *percloroestireno* y el *percloro*-p-*xilileno.* Mientras que sus correspondientes hidrocarburos, el estireno y el *p*-xilileno son inestables, sensibles al aire y sufren fácilmente polimerización, los compuestos en cuestión resisten sin alteración la acción de agentes de gran agresividad química, como por ejemplo el ácido sulfúrico concentrado caliente, no se oxidan, ni poseen la capacidad de polimerizarse.

He aquí pues dos resultados realmente excepcionales: la inercia química de los enlaces etilénicos, que normalmente son muy reaccionables, y la posividad en reacciones de sustitución electrófila aromática; concretamente, el percloroestireno no adiciona en su doble enlace ni sufre substituciones electrófilas de los cloros unidos a sus carbonos anulares.





EL IMPEDIMENTO ESTEREO, que ocurre en las reacciones químicas y se debe a la falta de espacio, queda aquí al descubierto en la molécula de perclorotolueno *(arriba)* que no puede acomodar sus ocho átomos de cloro. El tolueno *(abajo)* es el hidrocarburo aromático correspondiente.

ESQUEMA DE LOS INTENTOS FALLIDOS en la síntesis del perclorotolueno, a partir del tolueno *(compuesto único de la izquierda)*, **efectuados en dos fases: cloración exhaustiva del CH_3, primero y del anillo después** *(a)*, **o viceversa** *(b)*. **El perclorotolueno (u octaclorotolueno) es el cloruro de carbono alquilaromático más sencillo, de fórmula C_7Cl_8.**

Hemos aplicado este concepto de protección estérea molecular no sólo a la estabilización de estructuras normalmente poco estables, e incluso inestables, de la química orgánica convencional, sino asimismo para conferir una gran inercia a las que normalmente son estables.

Puesto que los percloroestilbenos, obtenidos por nosotros por reductocondensación del perclorotolueno, poseen una sorprendente inercia química y una gran estabilidad térmica (hasta 400°, por lo menos), decidimos obtener polímeros de estructura análoga a ellos, logrando así algunos (verbigracia, el PP-xineno) que llegan a resistir el ácido sulfúrico fumante (oleum) y el ácido nítrico fumante en caliente durante centenares de horas, así como temperaturas de hasta 600° C (rojo incipiente).

Tal resistencia no ha sido igualada aún por ningún polímero orgánico definido de síntesis del que el autor tenga noticia. Hemos obtenido, asimismo, percloropoliesteres y percloropoliamidas que presentan también una gran estabilidad química y térmica.

Es sabido que los acetilenos son químicamente todavía más activos que los etilenos, dando principalmente *reacciones de adición* (reciben tal denominación las que rinden una molécula a partir de dos o más). Algunos de ellos pueden originar incluso explosiones si no se manejan con el debido cuidado. En cambio, los percloroacetilenos obtenidos por nosotros son sorprendentemente estables, a pesar de que la hibridación digonal referida anteriormente brinda escasa protección estérea a su triple enlace. Por ejemplo hemos sintetizado el perclorofenilacetileno, que si bien puede reaccionar en condiciones relativamente vigorosas y se oligomeriza por la acción del calor, es francamente estable. Los derivados metálicos de los acetilenos (acetiluros), especialmente los de plata, pueden llegar a ser muy peligrosos por su gran sensibilidad al choque y al calor. Pues bien, el perclorofenilacetiluro argéntico es perfectamente estable, pudiendo aplicársele una llama o martillearlo vigorosamente sin peligro de explosión. Hay que puntualizar, sin embargo, que tanto el perclorofenilacetileno como el perclorofenilacetiluro de plata están protegidos por el par de átomos de cloro de las posiciones próximas *(orto)*.

Otra aplicación del concepto de protección estérea molecular desarrollada con éxito se centra en el dominio de los *radicales libres*.

En el año 1900, M. Gomberg, profesor de la Universidad de Michigan, anunció la obtención del primer radical libre. Se llamaba radical a lo que aquí hemos denominado sustituyente; por ejemplo, el radical metilo (CH_3)b Entonces los radicales eran concebibles sólo como parte integrante de las moléculas, sin existencia independiente- de la misma manera que una parte amputada del cuerpo humano no sigue viviendo separada de él.

Con el descubrimiento de Gomberg quedó demostrado que ciertos fragmentos de moléculas podían tener vida propia (radicales libres), aunque muy efímera, frecuentemente del orden de millonésimas de segundo. Actualmente, los radicales libres, que suelen formarse en bajísima proporción, se detectan mediante técnicas extraordinariamente sensibles, como la espectrometría de *reso-*

HIBRIDACIONES DEL CARBONO en la molécula orgánica. El carbono puede tener tres configuraciones: *a)* **la** *hibridación tetraédrica; b)* **la** *hibridación trigonal,* **y** *c)* **la** *hibridación digonal.* **En ellas el carbono se une a cuatro, tres y dos sustituyentes según los vértices de un tetraedro regular, un triángulo equilátero y un ángulo plano, respectivamente, como se muestra arriba.**

nancia paramagnética electrónica (resonancia de spin electrónico). Se trata, pues, de un tipo de materia formada por partículas de gran agresividad química que buscan activamente estabilizarse atacando la materia de su entorno.

Los radicales libres desempeñan un papel importantísimo en multitud de reacciones químicas, incluso de carácter industrial, como las polimerizaciones. Asimismo, son partículas esenciales en procesos bioquímicos, incluyendo los patológicos.

La característica fundamental de los radicales libres es la presencia de uno o más *electrones solitarios;* o sea electrones que no comparten su orbital molecular con otros, infringiendo en cierto modo la antigua regla de los dobletes de Lewis (asociación de los electrones de una molécula en pares). Tales electrones confieren a su molécula ciertas propiedades eléctricas y magnéticas, de las que la materia común carece, asociadas a una magnitud fundamental en física cuántica denominada *spin electrónico.* Puede decirse que el spin actúa como un imán elemental, pudiendo darse respecto a un campo magnético sólo dos orientaciones: a favor de él o en contra. Por consiguiente, se dice que el spin puede adoptar los valores de 1/2 ó -1/2.

Los radicales del elemento carbono, como el CH_3, sólo pueden producirse por anulación de una valencia del carbono tetravalente; o sea, pasando a trivalente. (Se han detectado, asimismo, especies de naturaleza esencialmente radicalaria de carbono divalente, que suelen llamarse, "carbenos", como el CH_2 o metileno.)

Los radicales libres del carbono suelen reaccionar velozmente con el oxígeno del aire, asociarse de dos en dos para anular su spin y formar especies estables (dímeros), y atacar activamente otras moléculas, incluso de reconocida pasividad química. Por ello, creímos no sólo interesante, sino muy importante, intentar protegerlos por percloración.

Ya en el desarrollo de la química de los percloroalquilbencenos, ejemplarizados aquí en el perclorotolueno y el percloro-*p*-xileno, se pusieron de manifiesto ciertos hechos que apuntaban hacia la existencia de un radical transitorio pero de notable estabilidad: el radical perclorobencilo ($C_6Cl_5CCl_2$), a pesar de que el correspondiente radical hidrocarbonado es una especie de existencia muy efímera. Hemos sintetizado hace poco aquel radical, resultando ser *estable a la temperatura ambiente,* encontrándose en equilibrio con su dímero, el perclorobibencilo.

En el radical perclorobencilo el carbono trivalente está protegido por los dos átomos de cloro unidos a él (*alfa*), así como por el grupo pentaclorofenilo: concretamente por los dos átomos de cloro de las posiciones *orto.* En la molécula de su dímero existe una gran repulsión entre los dos grupos perclorobencilo que la forman (tensión frontal) causando un debilitamiento importante del enlace entre los dos carbonos centrales que determina la existencia de tan insólito equilibrio. El radical perclorobencilo, sin embargo, no es aislable y es atacado rápidamente por el oxígeno atmosférico dando cloruro de perclorobenzoílo.

La substitución de los átomos de cloro de la posición *alfa* del radical perclorobencilo por grupos pentaclorofenilo, mucho más grandes que aquéllos, debiera brindar a los radicales resultantes una mayor protección de su carbono trivalente. Tales radicales han sido sintetizados por nosotros. El *radical perclorotrifenilmetilo* (o PTM), por ejemplo, posee efectivamente unas estabilidades química

ESTIRENO
ADICIONA Y SE POLIMERIZA

P-XILILENO
NO AISLABLE. SE POLIMERIZA
CON GRAN RAPIDEZ

PERCLOROESTIRENO
NO ADICIONA NI POLIMERIZA

PERCLORO-P-XILILENO
NO ADICIONA, OXIDA O POLIMERIZA

DOS EJEMPLOS DE MOLECULAS SENCILLAS, el percloroestireno y el percloro-*p*-xilileno, que ilustran perfectamente la inercia química de los perclorocompuestos etilénicos y aromáticos. Mientras que sus correspondientes hidrocarburos, el estireno y el p-xilileno son inestables, sensibles al aire y sufren fácilmente polimerización, los compuestos en cuestión resisten sin alteración la acción de agentes de gran agresividad química (por ejemplo, el ácido sulfúrico concentrado caliente), no se oxidan, ni son capaces de polimerizarse.

a

PERCLOROFENILACETILENO

PERCLOROFENILACETILURO ARGENTICO

b

RADICAL PERCLOROBENCILO

PERCLOROBIBENCILO

LA PROTECCION ESTEREA MOLECULAR se manifiesta incluso en la sorprendente estabilidad de especies como el perclorofenilacetileno y el perclorofenilacetiluro argéntico *(a)* y el radical perclorobencilo, que es estable a la temperatura ambiente y que se encuentra en equilibrio con su dímero, el perclorobibencilo *(b)*. El punto representa el electrón solitario.

...— CCl = CCl — C_6Cl_4 — CCl = CCl — C_6Cl_4 — CCl = CCl — C_6Cl_4 — ...

PP-XINENO

$$\ldots\text{-O-}C_mCl_n\text{-O-}\underset{\displaystyle O}{\overset{}{\underset{\|}{C}}}\text{-}C_rCl_s\text{-}\underset{\displaystyle O}{\underset{\|}{C}}\text{-O-}C_mCl_n\text{-O-}\underset{\displaystyle O}{\underset{\|}{C}}\text{-}C_rCl_s\text{-}\underset{\displaystyle O}{\underset{\|}{C}}\text{-}\ldots$$

PERCLOROPOLIESTER (ESTRUCTURA GENERAL)

DOS POLIMEROS de gran resistencia química y térmica: el PP-xineno, que es un cloruro de carbono macromolecular que resiste temperaturas de hasta 600ºC (rojo incipiente) y la acción de agentes muy agresivos (ácido sulfúrico, ácido nítrico, cloro, etc.), y los percloropoliesteres.

y térmica realmente inauditas. Su carbono trivalente (su auténtico talón de Aquiles), que reside exactamente en su centro geométrico, está perfectamente protegido del ataque exterior, hundido en su "bunker" de átomos de cloro.

El radical PTM, y muchos otros del mismo tipo, es fácilmente aislable en forma cristalina y en estado de pureza, según mediciones de susceptibilidad magnética, que muestran la existencia de un número de spines sensiblemente igual al número de Avogadro; es decir, $(6 \cdot 10^{23})$ spines. No se dimeriza en absoluto (enorme tensión frontal del hipotético dímero) y no es atacado por el oxígeno del aire. Mediciones de gran precisión realizadas con soluciones de radical PTM, en presencia de oxígeno, no logran detectar alteración alguna al cabo de meses. Extrapolando el pequeño error probable de las mediciones, se llega a la conclusión de que su vida media mínima es de cerca un siglo. En este punto debe destacarse que el radical de carbono trivalente más estable conocido en el momento en que se sintetizó el radical PTM era el 2(*p*-bifenilil)bisbifenilenilalilo, obtenido por Kuhn y Neugebauer, cuya vida media era de unas horas.

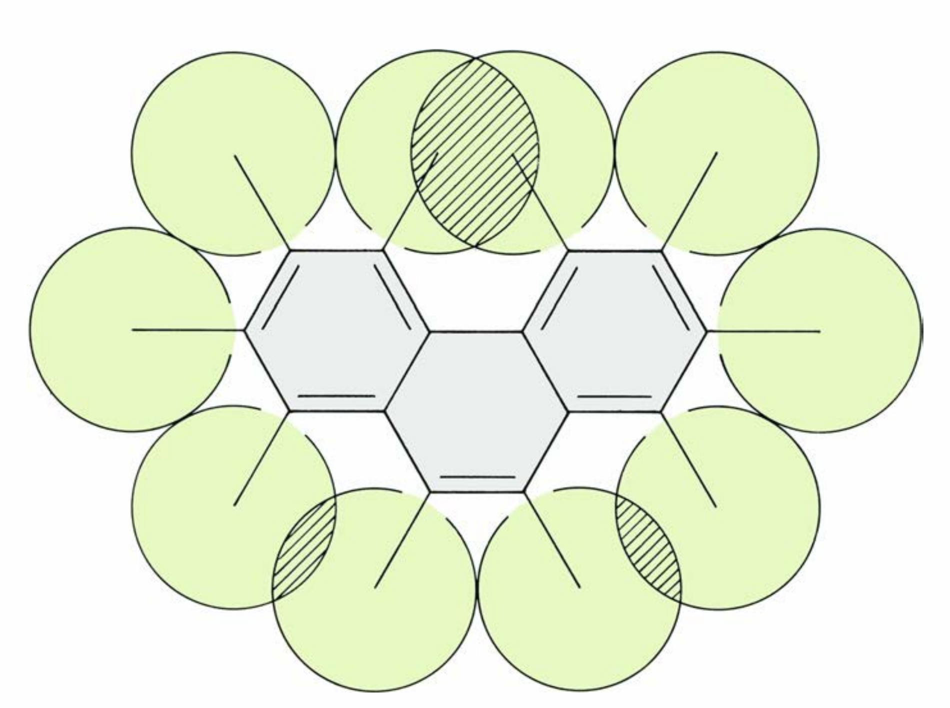
El radical PTM y sus análogos tampoco reaccionan con los reactivos típicos de los radicales libres, como son los halógenos, el óxido nítrico, la *p*-quinona, la hidroquinona, el tolueno hirviente, etc. Por añadidura, resisten la acción de agentes químicos tan corrosivos como el ácido sulfúrico concentrado, y puede alcanzar el aire temperaturas de hasta 300º C sin descomponerse. Por ello, puede decirse que estas especies de carbono trivalente poseen *una estabilidad térmica y una resistencia química superiores a las de la inmensa mayoría de compuestos orgánicos de carbono tetravalente de la química orgánica convencional.* Algunos científicos corresponsales nuestros han sugerido el nombre de *radicales libres inertes* para designarlos.

Con la síntesis de los rádicales libres inertes ha quedado abierto un nuevo campo de la química al estudio teórico y a la experimentación.

Es sabido que el *electrón* es una partícula muy pequeña; aproximadamente pesa unas 63.000 veces menos que un átomo de cloro. Por consiguiente, no era previsible que la protección que brindan los cloros a un radical libre inerte frente al ataque de especies químicas fuera también eficaz en procesos que implican la extracción o introducción de un electrón en ellos; es decir, en reacciones de oxidación o reducción, respectivamente. De acuerdo con ello, la acción de fuertes oxidantes o reductores sobre los radicales como el PTM los convierten en especies químicas con carga eléctrica positiva (iones carbonio) o negativa (carbaniones) de análoga estructura, que disfrutan asimismo de protección molecular por cloro. Por consiguiente, tales especies poseen una estabilidad relativamente elevada, lo que ha permitido aislarlas en forma de sales puras y cristalinas; por ejemplo, el hexacloroantimoniato de perclorotrifenilcarbonio y el perclorotrifenilmetiluro de tetraetilamonio.

Es digno de destacar que el perclorotrifenilcarbanión $(C_6Cl_5)_3C^{\ominus}$ es el único auténtico carbanión conocido que no es atacado por el agua en solución (no hidrolizable), lo que se debe indudablemente a su gran protección estérea.

Otro hecho importante, que muestra que los radicales químicamente inertes pueden ser activos en procesos de transferencia electrónica, es la construcción de una batería eléctrica orgánica experi-



REPRESENTACION BIDIMENSIONAL A ESCALA de la molécula del perclorofenantreno en su conformación plana. Las zonas sombreadas representan el solapamiento de las esferas de acción y son una medida de la importancia de las repulsiones entre los átomos de cloro.

mental por los Aerospace Research Laboratories (Ohio, Estados Unidos), basada en la actividad electrónica del radical PTM.

Los resultados anteriores muestran que los radicales libres inertes son intrínsecamente muy agresivos, pero resultan "inofensivos" en su jaula de cloro.

Un resultado sorprendente ha sido el descubrimiento de un nuevo mecanismo para una reacción de la química orgánica de gran importancia teórica y de aplicación en el proceso de refino de gasolinas: las *reacciones de transferencia de ion hidruro* (H^-) a ion carbonio. Merced a la protección por cloro ha sido posible demostrar que, contrariamente a la creencia general de que el proceso tiene lugar en una sola fase, ocurre en dos fases: primero, salto de un electrón al ion carbonio, dando el correspondiente radical, y luego transferencia de un átomo de hidrógeno al radical formado.

Este desdoblamiento en dos fases se debe a que mientras la transferencia del electrón no está impedida sí lo está la del átomo de hidrógeno y, por consiguiente, éste tiene lugar en el caso del ion perclorodifenilcarbonio (PDM^+) a velocidad medible.

En el caso del ion perclorotrifenilcarbonio (PTM^+), más impedido que el anterior, la transferencia se paraliza en el estadio intermedio, dando sólo radical PTM.

Si se substituye indirectamente en el radical PTM los átomos de cloro de las posiciones más alejadas del carbono (posiciones *para*) por substituyentes diversos se obtiene una gran variedad de radicales que tienen la cualidad de reaccionar en estos substituyentes sin afectar el carácter radicalario. Actualmente estamos desarrollando un programa de investigación cuyo objetivo es el marcaje de spin (spin tagging) o anclaje de radicales libres inertes a otras moléculas, transfiriendo a éstas tanto sus propiedades eléctricas y magnéticas como su inercia radicalaria, lo que puede permitir el estudio de las mismas mediante la técnica de resonancia de spin electrónico. Así se pretende conseguir el anclaje a proteínas, lípidos, etc. Un ejemplo reciente y sencillo de anclaje radical libre inerte es el caso del hidrobromuro de 4-(L-alanilglicil) aminotetradecaclorotrifenilmetilo.

a

$(C_6Cl_5)_2\dot{C}$ — NHOCCH$_2$NHOCCH$\overset{+}{N}$H$_3$ Br$^-$

b

RADICAL PERCLORODIFENILAMINILO

c

RADICAL PERCLORODIFENILNITROXIDO

UN EJEMPLO SENCILLO DE ANCLAJE de radical libre inerte es el hidrobromuro de 4-(L-alanilglicil)aminotetradecaclorotrifenilmetilo. A este tipo de marcaje de spin se atribuye cierta importancia en el estudio de la estructura de moléculas complejas, tanto de naturaleza bioquímica como industrial, y asimismo un considerable valor analítico *(a)*. Dos casos interesantes de extensión de protección estérea molecular por cloros a radicales de otros elementos químicos (nitrógeno) son la obtención en forma cristalina de los radicales perclorodifenilaminilo *(b)* y el perclorodifenilnitróxido *(c)*, dotados éstos de sorprendente estabilidad.

Hemos sintetizado, asimismo, birradicales inertes que muestran propiedades de interés relacionadas con el spin electrónico y sus acoplamientos con núcleos atómicos magnéticamente activos, y con los procesos de absorción electrónica.

Una importante cuestión no investigada aún se refiere a las propiedades bioquímicas e incluso farmacológicas de los radicales libres inertes.

Una aplicación interesante estriba en la conversión de radicales libres inestables en radicales libres inertes por captura de aquéllos mediante reactivos perclorados adecuados, habiéndose logrado ya resultados valiosos.

Estamos extendiendo actualmente la aplicación del concepto de protección estérea molecular por cloros a radicales de otros elementos químicos, particularmente de nitrógeno, habiendo obtenido en forma cristalina el *radical perclorodi-*

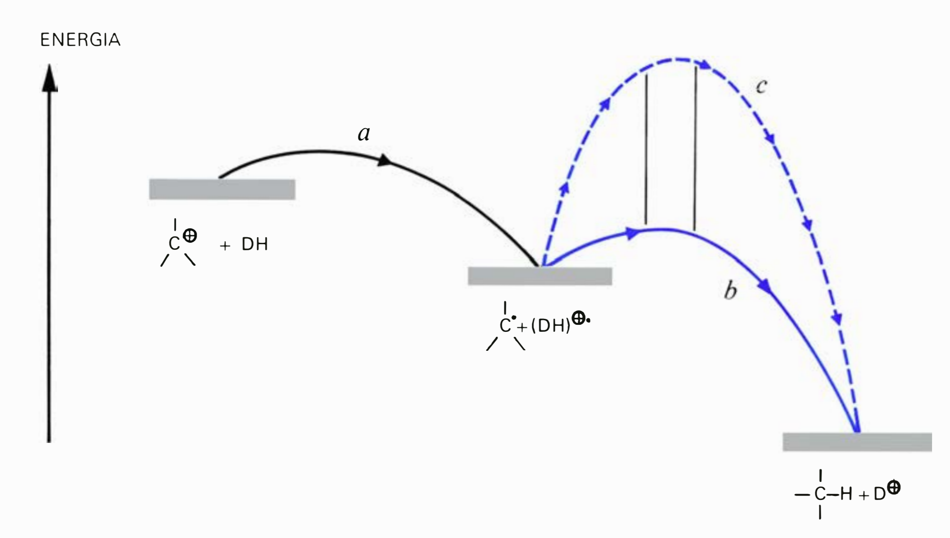


DIAGRAMA ENERGETICO del proceso de transferencia de ion hidruro, en donde la especie DH es el dador de ion hidruro. La primera fase es la transferencia de un electrón *(a)*. La alta energía de activación de la segunda fase del proceso *(b)*, cuando existe impedimento estéreo, posibilita la detección del radical intermedio perclorodifenilmetilo (PDM) por resonancia de spin electrónico. En el ion perclorotrifenilcarbonio (PTM) el impedimento es tan elevado que la segunda fase de la reacción *(c)* no tiene lugar. A mayor altura de la segunda, mayor impedimento.

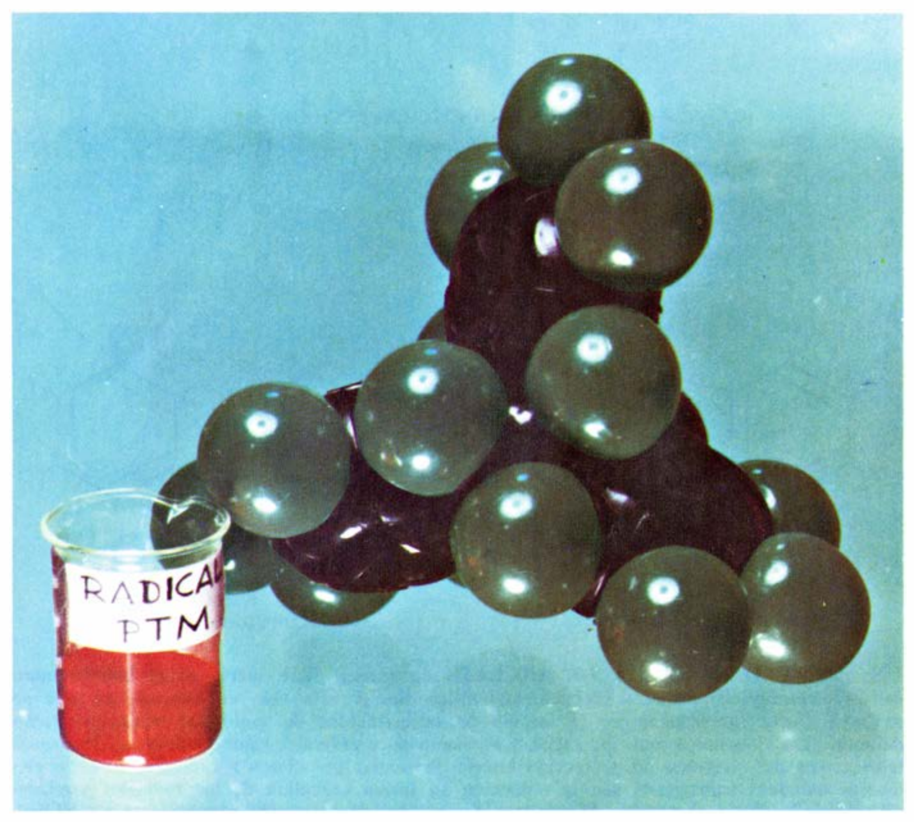
fenilaminilo (radical de nitrógeno bivalente) y el *radical perclorodifenilnitróxido*.

Los radicales nitróxido, que pueden ser considerablemente estables al aire, suelen reaccionar sin embargo con ácidos bajo condiciones relativamente suaves, y con los halógenos, por ejemplo. Contrariamente, el radical perclorodifenilnitróxido resiste el ácido trifluoroacético y el cloro molecular.

Se sabe que los compuestos halogenados se distinguen especialmente por su actividad biológica, por lo que se emplean profusamente como insecticidas, germicidas, fungicidas, herbicidas, etc., así como en medicina. Se sabe, asimismo, que algunos compuestos estrechamente emparentados con los aquí descritos son activos contra la esquistosomiasis y la malaria. Algunos derivados fenólicos preparados por nosotros por substitución nucleófila en compuestos percloroaromáticos muestran actividad biológica. Alguno de ellos es, como bactericida y fungicida, incluso algo más activo que el hexaclorofeno.

Otras investigaciones en desarrollo se refieren a nuevos polímeros de gran resistencia química y térmica adaptados a ciertas aplicaciones (aislantes eléctricos, semiconductores orgánicos, adhesivos, lubricantes, cromatografía de gases a alta temperatura, etc.).

Entre los resultados de naturaleza teórica destaca especialmente una correlación muy útil entre la estructura de las moléculas derivadas del *benceno* y su *espectro ultravioleta* de absorción, que a pesar de haberse hallado en el dominio de los compuestos perclorados y aplicado a la elucidación de sus estructuras, se extiende asimismo al campo de los compuestos orgánicos convencionales. Mediante tal correlación, llamada "ley lineal", hemos podido averiguar, por ejemplo, que importantes estudios acerca de la asignación de bandas en los espectros de absorción de las anilinas eran parcialmente inexactos, conclusión corroborada posteriormente por otros autores partiendo de estudios y consideraciones independientes.



FORMULA DEL RADICAL PTM

EL RADICAL PTM es el paradigma por excelencia de un radical libre inerte. Arriba, se representa su modelo a escala (adjunto a una muestra substancial del mismo) y, abajo, su fórmula estructural. El carbono trivalente, situado en el centro de la molécula, está bien protegido por seis cloros (sólo tres de estos cloros son visibles) y tres anillos bencénicos.

ION CARBONIO + ION HIDRURO (UNIDO AL DADOR) → RADICAL LIBRE + H· | C· + ATOMO DE HIDROGENO (UNIDO AL DADOR) → PRODUCTO (—C—H)

LA TRANSFERENCIA DE ION HIDRURO se da entre un dador, el cicloheptatrieno por ejemplo, y un ion perclorocarbonio, verbigracia el perclorodifenilcarbonio. El par de puntos sobre el ion hidruro representa su par de electrones que se escinde en el proceso. Merced a la protección por cloro, ha sido posible demostrar que, contrariamente a la creencia general, el proceso ocurre en dos fases: primero, salto de un electrón al ion carbonio, dando el correspondiente radical y, luego, transferencia de un átomo de hidrógeno a este radical en cuestión.

¿Ha leído Ud.
estos dos números?



Nuestros suscriptores, sí

Las distintas ediciones internacionales de SCIENTIFIC AMERICAN son leídas por casi 800.000 personas. La inmensa mayoría de ellas tienen asegurada la cómoda recepción mensual de la revista en sus domicilios. También con INVESTIGACION Y CIENCIA se está produciendo este fenómeno. Nuestros lectores desean asegurarse la recepción de sus ejemplares y obtener gratuitamente el número extraordinario anual. INVESTIGACION Y CIENCIA es, por sus características, una publicación especialmente idónea para que el lector se convierta, desde el primer momento, en suscriptor. En sus manos tiene Ud. la mejor demostración: la propia revista.

LA TARJETA DE SUSCRIPCION ENCARTADA EN ESTE EJEMPLAR LE FACILITA LA COMODA RECEPCION MENSUAL DE SU REVISTA

Juegos matemáticos

Problemas de combinatoria, viejos unos, otros nuevos, tratados mediante ordenador

Martin Gardner

La "revolución de la combinatoria", en matemáticas, está todavía en auge, ya que los libros y los artículos sobre combinatoria continúan proliferando. El ordenador electrónico está contribuyendo ciertamente a dicha revolución, mediante su poder de analizar problemas de combinatoria demasiado complejos para ser abordados de cualquier otra manera. Otra fuerza estimulante es la creciente aplicación de la teoría combinatoria a la ciencia y a la tecnología, particularmente en física de partículas y en biología molecular. A gran escala, puede que el universo sea una colección de continuos de los que pueda dar razón el cálculo, pero al nivel microscópico es una mezcla de elementos discretos que dan saltos misteriosos y tienen curiosas propiedades combinatorias. En algunas teorías modernas, incluso el tiempo y el espacio están cuantificados.

Cientos de enigmas fascinantes de combinatoria, algunos viejos, otros nuevos, están recibiendo la atención de los matemáticos serios. Echemos primeramente un vistazo a un divertido y excelente rompecabezas inventado por Sam Loyd, que un programa de ordenador acaba de invalidar; después continuaremos con varios problemas de combinatoria que no tienen una solución general a la vista, pero presentan aspectos sugestivos en sus niveles más sencillos.

El mayor libro de rompecabezas que nunca se haya publicado en Estados Unidos es un gran volumen, encuadernado en tela verde claro, que lleva escrito en sus tapas: *Sam Loyd's Cyclopedia of 5000 Puzzles, Tricks and Conundrums, with Answers. (Enciclopedia de Sam Loyd con 5000 rompecabezas, trucos, acertijos y sus respuestas).* El precio de cubierta es de 5 dólares, pero en la actualidad es difícil encontrar un ejemplar por menos de 25 dólares. No se sabe casi nada sobre la historia de su publicación. Aunque está profusamente ilustrado, no se da el nombre de ninguno de los artistas. Los ejemplares fueron editados por Morningside Press y por Lamb Publishing Company, ambas de Nueva York, pero nadie sabe cuál es la primera de las dos ediciones.

Todos los ejemplares están fechados en 1914. Puesto que esta fecha es tres años posterior a la muerte de Loyd, se supuso que su hijo, que tenía el mismo nombre que su padre, fue quien seleccionó los rompecabezas de periódicos y revistas viejos, reuniéndolos apresuradamente en un gran volumen. (El libro está salpicado de omisiones, faltas y errores de imprenta.) Se descubrió más tarde que el viejo Loyd había editado y publicado una revista trimestral llamada *"Our Puzzle Magazine"*, que apareció por primera vez en junio de 1907. La *Cyclopedia* es simplemente una reedición de esta revista hecha a partir de las planchas de sus páginas.

El rompecabezas *"Back from the Klondike" (véase la ilustración de arriba en la página siguiente)* aparece en el segundo número (octubre de 1907) de la revista de Loyd y está reeditado en la página 106 de la *Cyclopedia.*

"Se empieza desde el corazón que hay en el centro", dice Loyd, y se dan tres pasos en línea recta en una de las ocho direcciones posibles: norte, sur, este, oeste, nordeste, noroeste, sudeste, suroeste. Cuando se han dado los tres pasos en línea recta, se alcanza un cuadrado que tiene un número escrito, número que indica los pasos que habrán de darse en el segundo día de viaje, en línea recta y en una cualquiera de las ocho direcciones. Al llegar a este nuevo punto, se volverá a emprender la marcha de acuerdo con el número indicado y se continuará adelante siguiendo las indicaciones de los números a los que se va llegando, hasta alcanzar un cuadrado que tenga un número que lleve justamente un paso más allá del borde. Entonces se supone que se está fuera del bosque y ya puede usted vocear todo lo que quiera, porque habrá resuelto el rompecabezas.

La solución de Loyd la indica la línea coloreada. El camino consiste simplemente en ir arriba y abajo de la diagonal principal hasta conseguir un número que haga posible lo que Loyd, en su respuesta, llama un "descubrimiento audaz vía S. E. hacia la libertad". La solución, según dice Loyd, es única. "Los que no lograron encontrarla descubrieron en seguida que un paso mal dado en cualquiera de las fases del juego le arroja a uno dentro de un torbellino del cual no se puede salir".

Pero el viejo maestro estaba equivocado. A principios de este año, Penelope J. Greene (graduada en sociología por la Universidad de Washington), R. Duncan Mitchell (graduado en económicas) por la misma Universidad y Horace A. Greene (profesor de ciencias en la Universidad Roosevelt de Chicago), hicieron un ataque conjunto al problema mediante un programa escrito en Fortran por Mitchell. En pocos segundos el programa encontró la solución de Loyd juntamente con cientos de otras. De hecho, se puede dejar el centro 3 siguiendo una cualquiera de las ocho direcciones y escapar del bosque. El comienzo de otra ruta posible se muestra en la ilustración inferior de la página siguiente. Todas estas rutas conducen finalmente al 4, enlace clave sobre la diagonal, desde el que el camino se completa como en la solución de Loyd.

Los tres programadores notaron algo anormal en las respuestas posibles. Una célula, y sólo una, es común a todas estas otras rutas, célula que, además, no interviene en la solución de Loyd. Ellos la llaman el "inoportuno 2" (en la ilustración aparece en un círculo). ¿Podría tratarse de un error del artista? Parece probable, ya que cuando los programadores sustituyeron el 2 por otros dígitos, el computador solamente encontró un dígito distinto de cero tal que todos los demás posibles caminos quedaban eliminados.

Esta es la grata tarea –realmente es bastante fácil– que ofrezco a los lectores antes de darles a conocer la respuesta el mes próximo: corrijan el rompecabezas Klondike cambiando el inoportuno 2 por el único dígito distinto de cero que devuelve la unicidad a la solución de Loyd. Llamo única a la solución de Loyd, descartando las "excursiones laterales"

en las que lo único que se hace es dejar una célula para volver a ella de nuevo antes de seguir adelante o en las que se deja la célula A para ir a la B por un camino lateral cuando se podría ir directamente de A a B.

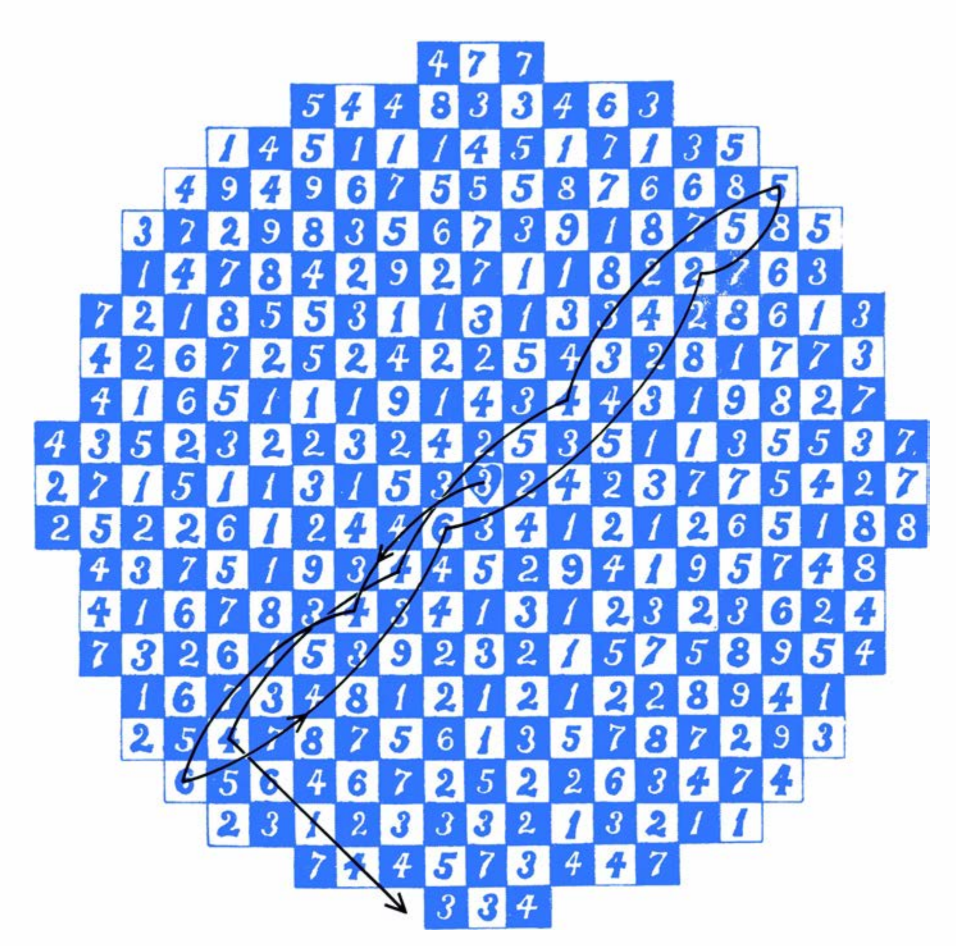
El origen del juego de damas chino, juego de tablero que es popular en todo el mundo aproximadamente desde 1920, es oscuro. Parece que fue puesto a la venta por primera vez bajo este nombre comercial en Estados Unidos no tiene nada que ver con China y casi no tiene parecido con el juego de damas, pero el nombre ha persistido.

Los agujeros del tablero standard *(véase la ilustración de la pág. 102)* están numerados para poder registrar los movimientos y se añaden bordes de color para perfilar todos los tableros estrellados más pequeños. Si juegan dos personas se colocan 10 bolas de un color en el "área" triangular que forman los agujeros del 1 al 10, y 10 bolas de otro color van en los agujeros 112 a 121. Las reglas del juego son parecidas a las del Halma, juego popular británico, que probablemente sugirió la variante representada por el juego de damas chino.

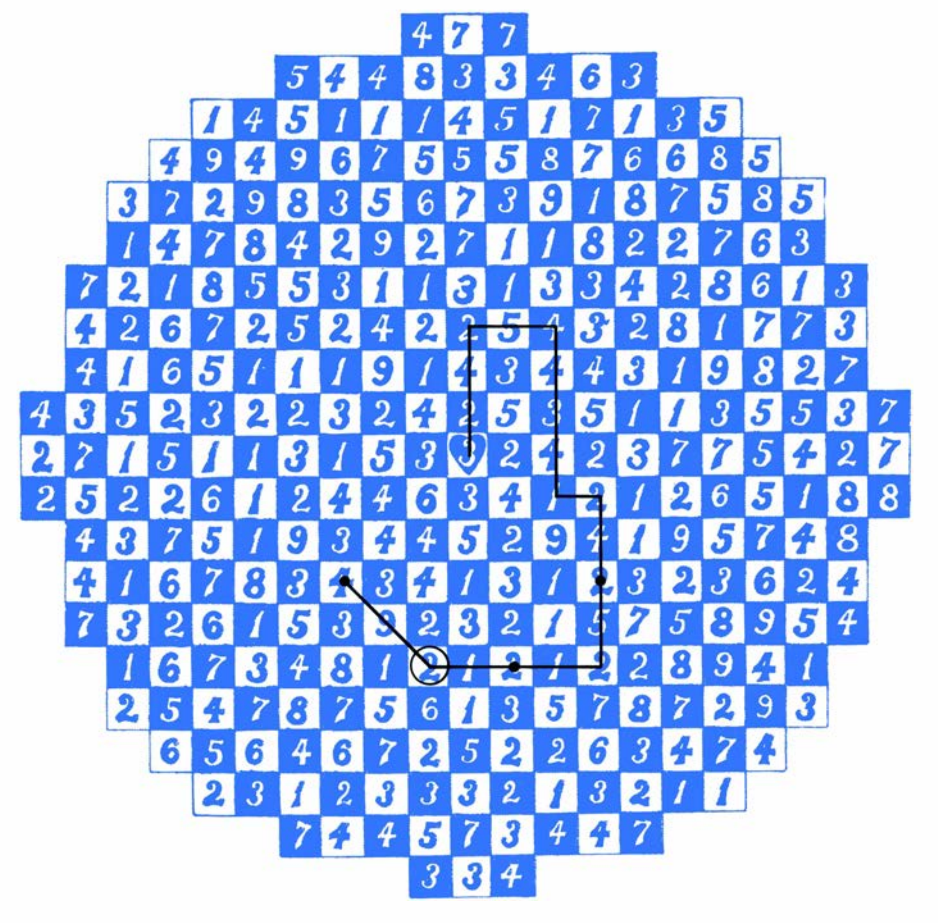
Están permitidos dos tipos de movimientos. Un "paso" es un movimiento hecho en una de las seis direcciones hacia un agujero adyacente que esté vacante. Un "salto" es un movimiento sobre una bola adyacente, que también puede ser en una cualquiera de las seis direcciones, hacia un agujero vacante al otro lado. El "salto" es igual a un salto de las damas, con la excepción de que la pieza saltada no se quita y puede ser tanto una pieza propia como una del contrario. Una pieza puede continuar saltando tanto como sea posible o puede pararse la cadena en cualquier punto. Los saltos no son nunca obligatorios. La combinación de "pasos" y "saltos" está prohibida. El ganador es el primer jugador que ocupe todos los agujeros del "área" de salida del jugador contrario.

El juego no puede, obviamente, jugarse sobre la "estrella" más pequeña, que corresponde a una "estrella" degenerada de un solo agujero. En la estrella que seguiría en tamaño a la anterior, cada jugador tiene una bola y el juego lo gana el jugador que mueve primero. La complejidad del juego aumenta enormemente cuando se utiliza un tablero en el que cada jugador necesita tres bolas. Creo que en este caso no se sabe qué jugador tiene las de ganar.

Sobre un tablero de este último tipo, el juego es bastante más interesante. Para evitar que un jugador fuerce un empate al mantener una bola perpetuamente



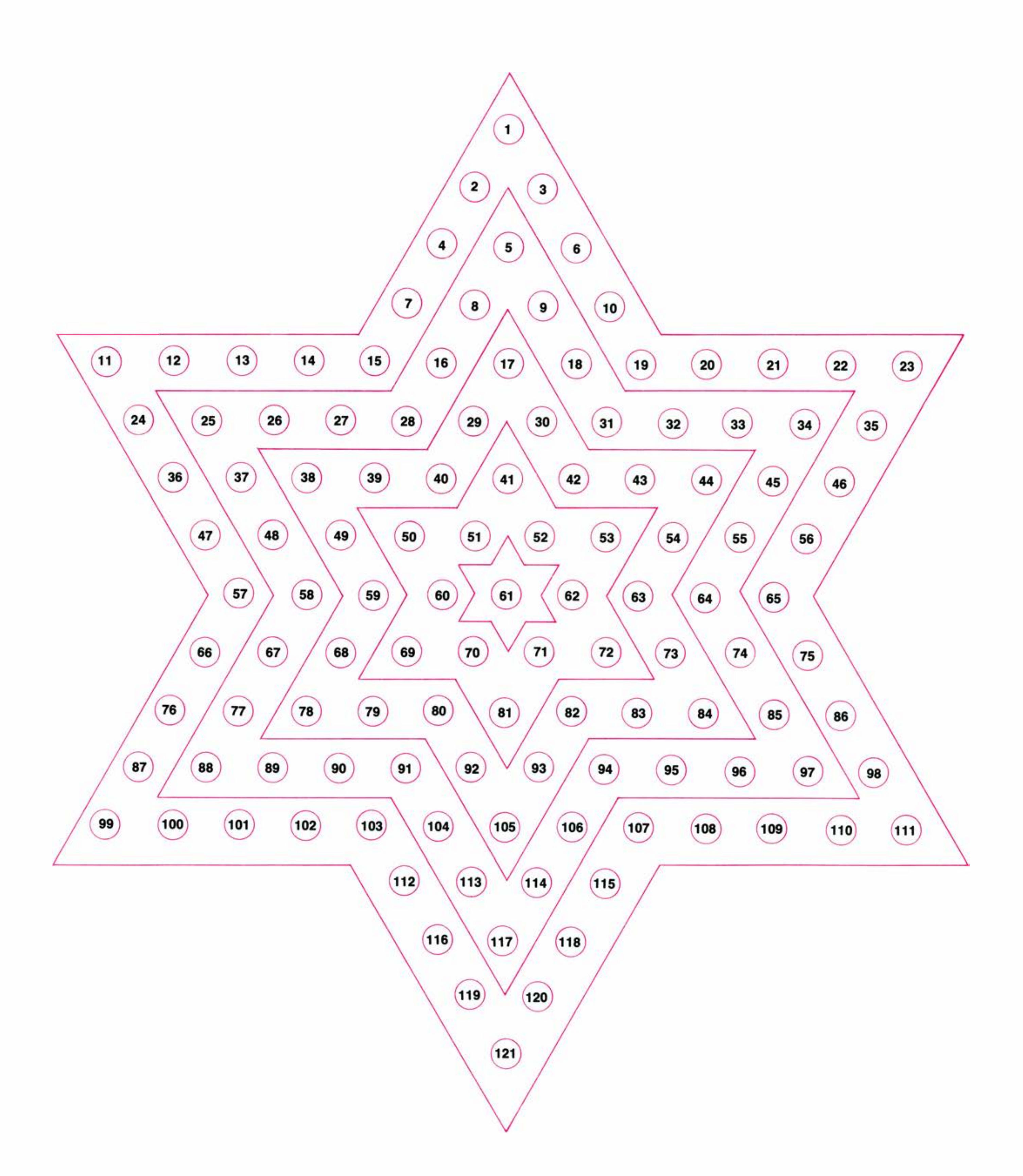
La solución de Sam Loyd a su rompecabezas Klondike



Un posible camino a través del "inoportuno 2"

en su "área" de salida, debe añadirse una regla de juego más: si una ficha puede salir de su "área" saltando a una ficha enemiga, debe hacerlo. Una vez esté fuera de su "área" no podrá volver: sin embargo, está permitido pasar a través del "área" mediante saltos sucesivos.

Considérese en cada tablero estrella el siguiente problema aislado. Se colocan bolas de un color en el "área" inferior, en posición de salida. ¿Cuál es el mínimo número de movimientos (contando una cadena de saltos como un solo movimiento) necesarios para llevar todas las bolas al "area" superior?



Diversos tableros del juego de damas chino

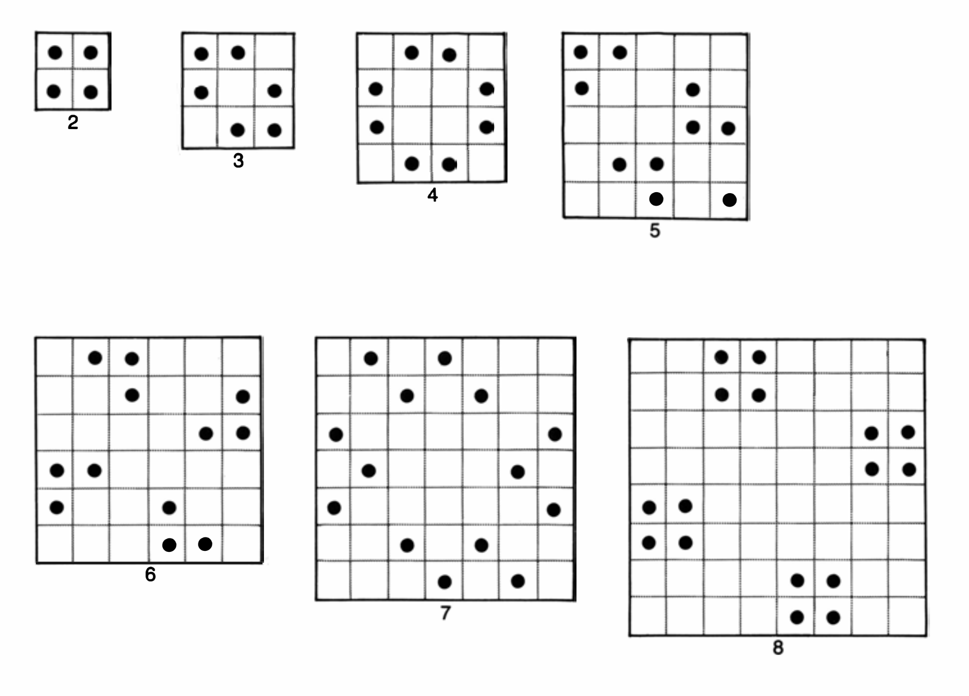
El problema no tiene sentido sobre la estrella más pequeña, y es trivial sobre la estrella de 13 agujeros. Para la estrella con un "área" de tres bolas, se puede lograr con 11 movimientos. Por ejemplo: 92-81, 81-71, 105-82-61, 61-52, 93-82, 82-61-42, 42-30, 71-61, 61-42-17, 52-41, 41-29.

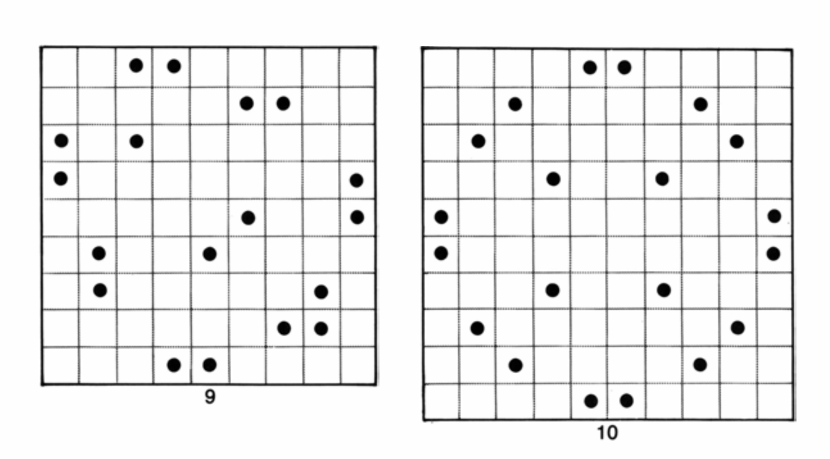
Para una estrella con seis bolas, la mejor solución que he sido capaz de obtener es de 18 movimientos. Empezando con 113-93 hay muchas maneras de construir en nueve movimientos una "escalera" vertical de bolas sobre los agujeros 9, 30, 52, 71, 93 y 114. Nueve movimientos más, cambiando el orden de la jugada anterior, colocan las seis bolas en el "área" superior.

Vayamos ahora a nuestro problema principal: ¿cuál es el número mínimo de movimientos necesarios para trasladar 10 bolas de un "área" a la otra sobre el tablero de damas chino? La primera vez que tuve noticia de este problema fue en 1961 por medio de Octave Levenspiel, profesor de Ingeniería química de la Universidad estatal de Oregón; él decía que lo máximo que había logrado fue hacerlo en 31 movimientos, pero que su madre, "la solucionadora de los rompecabezas de la familia, la campeona de ajedrez y de bridge", lo había resuelto una vez con 28, si bien no había podido recordar los movimientos. *Ibidem,* revista canadiense sobre magia (ya desaparecida), publicó en agosto de 1969 una demostración de que 29 era el mínimo. Luego en 1971, Levenspiel, intentando reconstruir la solución de su madre en 28 movimientos, dio el golpe, logrando una solución espectacular con 27 movimientos. Harry O. Davis, de Portland, Ore., cree tener una demostración de que 27 es el mínimo, pero la proposición no ha sido confirmada todavía.

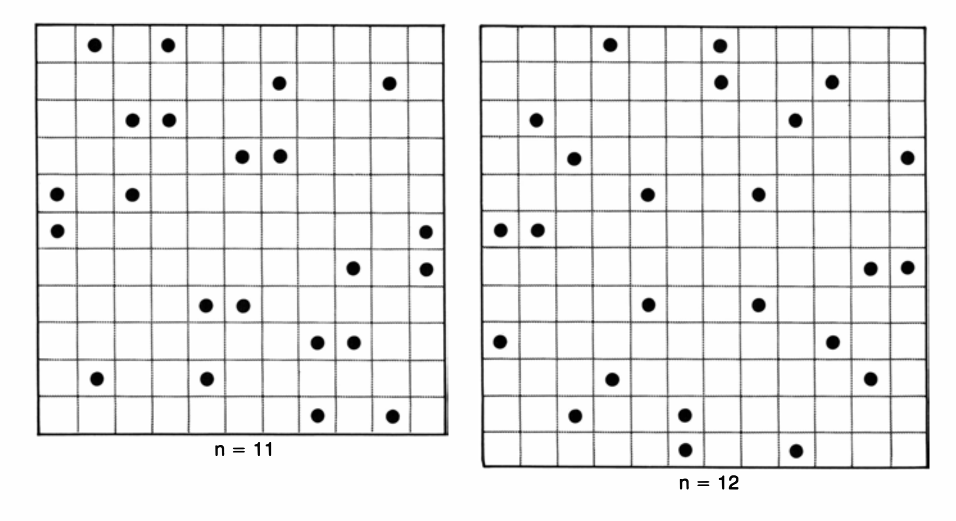
¿Podrá encontrar el lector una solución en 27 movimientos? Me alegraré mucho de saber que alguien logre hacerlo con menos movimientos o mejore alguno de los mínimos que he dado para tableros más pequeños. Los juegos de damas chinos se venden a veces con un tablero más grande que requiere 25 bolas para cada "área". Para un tablero así, la mejor solución que tengo en mi archivo, y que me fue enviada en 1974 por Min-Wen Du de Taiwan, es de 35 movimientos.

En el *Tribune* de Londres del 7 de noviembre de 1906, Henry Ernest Dudeney, que vendría a ser el Loyd inglés, publicó el problema de colocar 16 peones en un tablero de ajedrez de manera que no haya tres que estén en línea recta en ninguna dirección. (Este problema apa-





Algunas soluciones para el problema de no-tres en raya



Recientes soluciones al problema

ció más tarde con el n.º 317 en su *"Amusements in Mathematics".*) "Línea" significa cualquier línea recta y no solamente las ortogonales y diagonales. Los peones representan puntos en el centro de las células. Desde entonces, dicho problema, generalizado, ha sido el tema de varias publicaciones técnicas, bajo el nombre de "no-tres en raya".

El principal problema, que está todavía sin resolver, está en responder a la siguiente pregunta: sobre un tablero de damas cuadrado, con n células por lado (n mayor que 1), ¿es siempre posible colocar $2n$ fichas de manera que no estén tres en línea recta? El viejo principio de la "casilla" demuestra que $2n$ no puede ser rebasado. Ninguna fila o columna puede tener más de dos fichas, y ya que con $2n$ fichas es preciso colocar 2 en cada fila o columna, una ficha más aseguraría el tener que colocar tres fichas en una fila y otras tres en una columna. Un motivo menos trivial (descubierto por R. R. Hall, T. H. Jackson, A. Sudbery y K. Wild en su artículo "Some Advances in the No-Three-in-Line Problem" (en el *Journal of Combinatorial Theory,* serie A, vol. 18, mayo 1975, pág. 336-341) demuestra que pueden colocarse siempre al menos n fichas. Para grandes tableros, estos autores muestran que puede llegarse bastante cerca de $3n/2$ fichas.

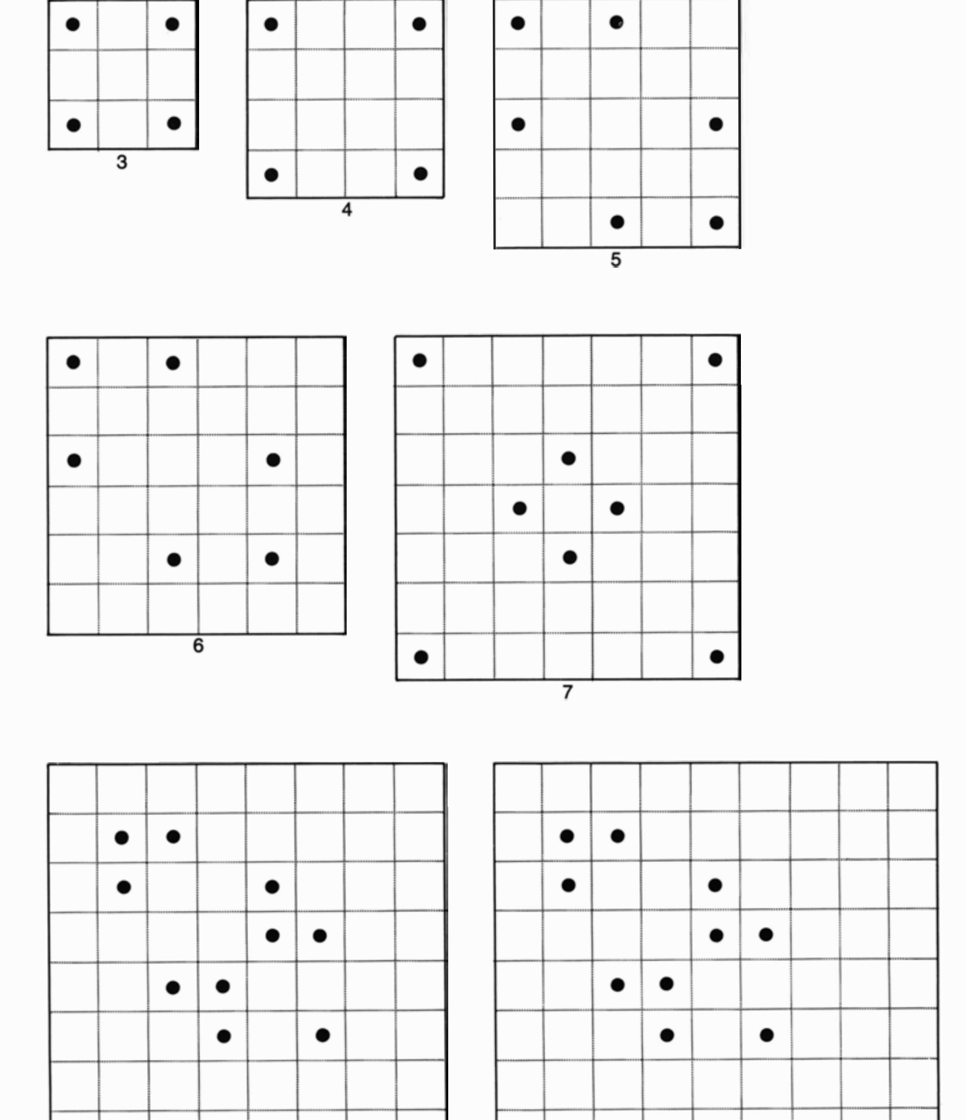
Michael A. Adena, Dreck A. Holton y Patrick A. Kelly, en su artículo "Some Thoughts on the No-Three-in-Line Problem" (publicado en *"Combinatorial Mathematics: Proceedings of the Second Australian Conference;* editado por Holton, Vol 403 de *"Lecture Notes in Mathematics",* Springer-Verlag, Berlín, 1974), dieron a conocer programas de ordenador que habían encontrado todas las soluciones distintas desde n igual a 2 hasta n igual a 10. Rotaciones y simetrías están excluidas. El número de soluciones es, respectivamente, 1, 1, 4, 5º 11, 22, 57, 51 y 1567. La ilustración superior *(pág. 103)* da un ejemplo para cada valor de n de 2 a 10. Nótese la asombrosa simplicidad y la simetría de la solución para un orden igual a 8.

En la época de este trabajo no se conocía ninguna solución para $n = 11$; su solución para $n = 12$ no resultó válida puesto que no se habían dado cuenta de que había dos filas de tres en línea.

Soluciones para $n = 11$ y $n = 12$ se encontraron en 1975 y sus descubridores fueron D. Craggs y R. Hughes-Jones de la Universidad de Kent, publicadas en el *"Journal of Combinatorial Theory"* (Serie A, v. 20, mayo 1976, pp. 363-364). *Véase la ilustración de la página anterior, derecha, abajo.* D. Graggs y R. Hugghes-Jones encontraron cinco soluciones más para $n = 11$ y otras tres para $n = 12$. El número total de soluciones para estos dos valores de n no se conoce y nadie ha construido una solución para cuadrados de orden superior a 12. Richard K. Guy y Patrick A. Kelly en su artículo "The No-Three-in-Line Problem" (*Canadian Mathematical Bulletin;* vol. 11, n.º 4, 1968, pág. 527-531) dieron razones en favor de su afirmación de que el número de órdenes con $2n$ soluciones es finito. No se sabe cuál es el tablero más pequeño en el que una solución de $2n$ sea imposible.

Si limitamos la definición de "línea" solamente a las filas, columnas y diagonales, el problema está resuelto. El máximo no puede ser superior a $2n$ y $2n$ es posible en todos aquellos tableros de orden superior a 1. Para n mayor que tres el problema se resuelve superponiendo dos soluciones cualesquiera del problema clásico "reinas no-atacantes". (Es el problema de colocar n reinas de manera que ninguna ataque a la otra.) Siempre es posible hacerlo con un par de soluciones que ponen las $2n$ reinas en $2n$ células.

En lugar de preguntar por el máximo número de fichas que pueden ponerse en un tablero de orden n, de manera que no estén tras en línea recta, preguntémonos por el mínimo número de las que pueden colocarse de manera que, añadiendo una ficha más en alguna célula vacante, se



Soluciones para el problema mínimo

obtengan tres en línea recta. Este interesante problema no ha recibido todavía la atención seria de los expertos.

Si la palabra "línea" se toma en su sentido más amplio (una línea recta con cualquier orientación), el problema podría ser muy difícil. Adena, Holton y Kelly lo mencionan brevemente en su artículo y dan como mejores resultados los siguientes (que obtuvieron "a mano") para *n* igual a 3 hasta *n* igual a 10: 4, 4, 6, 6, 8, 8, 12, y 12; por lo que yo sé, se desconoce como continúa esta sorprendente serie.

El problema no está tampoco resuelto si la palabra "línea" se restringe a las filas, columnas y diagonales. En otras palabras, ¿cuál es el número mínimo de peones que pueden colocarse en un tablero de lado *n* de manera que no pueda añadirse ningún peón más sin conseguir tres en una fila, una columna o en una diagonal? Otra manera de abordar este problema es considerarlo como un juego en el que dos jugadores colocan alternativamente fichas en un tablero cuadrado hasta que un jugador pierde, al verse forzado a colocar tres en una fila, una columna, o en una diagonal. ¿Cuánto puede durar un juego de este tipo? En la *Technology Review* (junio 1967) y en la excelente sección de Allan J. Gottlieb: *"Puzzle Corner"*, se propuso tal cuestión respecto de un tablero de órden 8. Esta es la única publicación que yo conozco de un problema de este tipo.

La ilustración inferior de la página opuesta muestra los mejores resultados que he encontrado para *n* igual a tres hasta *n* igual a 9. No he intentado encontrar todas las soluciones para cada uno, sino tan sólo para los tableros más pequeños. Hay docenas de soluciones para $n = 8$, muchas con simetría bilateral y otras con doble simetría. (Si 10 es el mínimo para el cuadrado de orden 8, la simetría cuadruple está excluida, ya que 10 no es divisible por 4.) ¿Podría el lector construir, antes de la respuesta del mes próximo, un único modelo de 12 fichas que diera simultáneamente la solución para los órdenes, 10, 11 y 12? No puedo demostrar que 12 sea el mínimo para estos tres valores, pero sospecho firmemente que es así.

El problema del mes pasado fue encontrar la respuesta ganadora al movimiento de un primer jugador en un juego llamado Cutcake. El pastel es un rectángulo de 4×7. Si el primer jugador rompe el pastel verticalmente en dos partes, un cuadrado de 4×4 y un rectángulo de 4×3, la única respuesta que permite ganar es romper el rectángulo de 4×3 en dos rectángulos de 2×3.

Taller y laboratorio

Unos dispositivos mecánicos que dibujan armonigramas

C. L. Stong

Las figuras ornamentales incluidas en este artículo pueden interpretarse de muchas maneras: como música representada, como el flujo de los electrones en un aparato de radio, o, simplemente, como figuras bonitas. En todos los casos se trata de la superposición de dos o más movimientos armónicos, en todo semejantes a los que muestra un péndulo al oscilar. Debido a ello, tales figuras se llaman *armonigramas.* Atrajeron el interés general en 1857 cuando el físico francés Jules Antoine Lissajous presentó su armonígrafo, aparato destinado a crear armonigramas.

En el aparato de Lissajous se empleaba la luz para trazar las curvas. Para ello Lisajous pegó pequeños espejos a un par de diapasones, que vibraban con frecuencias distintas en planos separados. Un rayo de luz reflejado por los dos espejos incidía en una pantalla próxima, en donde las vibraciones generaban una serie de curvas que se entrelazaban y cambiaban lentamente. La demostración fascinó a los auditorios, de modo que el diseño y manejo de armonígrafos se convirtió en un pasatiempo muy generalizado.

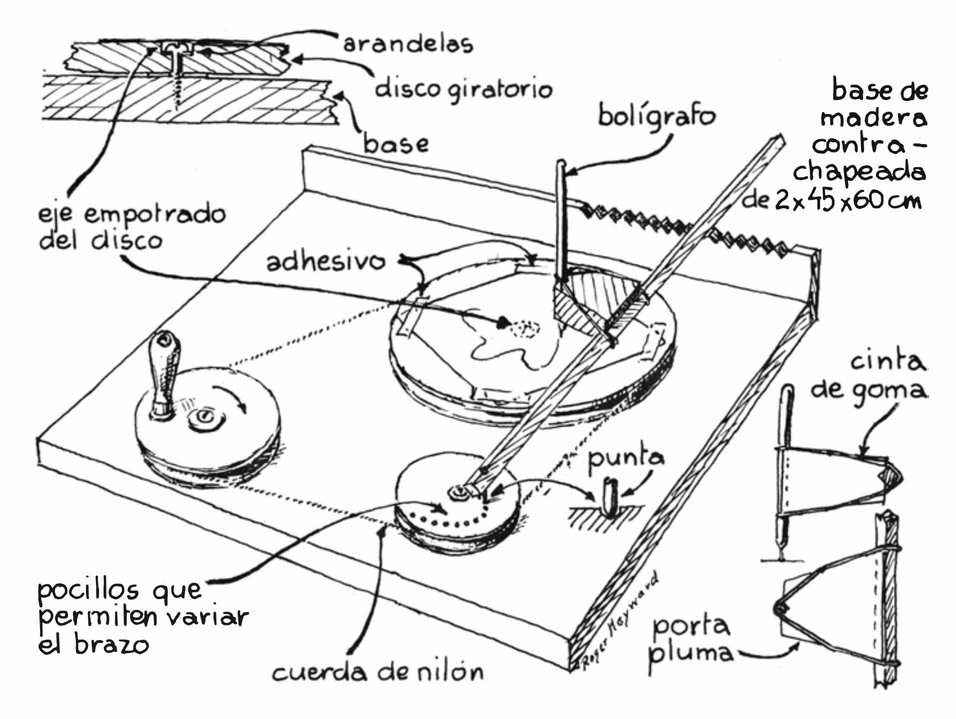
El pasatiempo continúa teniendo adeptos. Hace poco por ejemplo, Katherine O. Reed de Concordia, Massachusetts, Estados Unidos, junto con su padre, Thomas B. Reed, químico del laboratorio Lincoln del Instituto de Tecnología de Massachusetts, construyeron un armonígrafo para un concurso científico de su escuela superior. Cuando éste hubo terminado, el padre se encontró inmerso en el centenario tema. Como consecuencia de ello, los Reed han diseñado una sucesión de armonígrafos para crear figuras cada vez más atractivas. Su máquina más perfeccionada es complicada y pesa más de 30 kilogramos.

Así es como explican de qué se trata. Los armonigramas pueden producirse por diversos dispositivos que tienen ciertos elementos en común. Uno de los esenciales es un inscriptor de algún tipo, que puede consistir en pluma y papel, en un rayo de luz que actúa sobre papel fotográfico, en un estilete que graba sobre un vidrio ahumado, en una chispa eléctrica que oscurece un papel conductor, o en un rayo de electrones cuyos puntos de impacto sobre una pantalla son puestos de manifiesto por una sustancia fosforescente. La "pluma", cualquiera que sea, debe ser movida por un mecanismo oscilante.

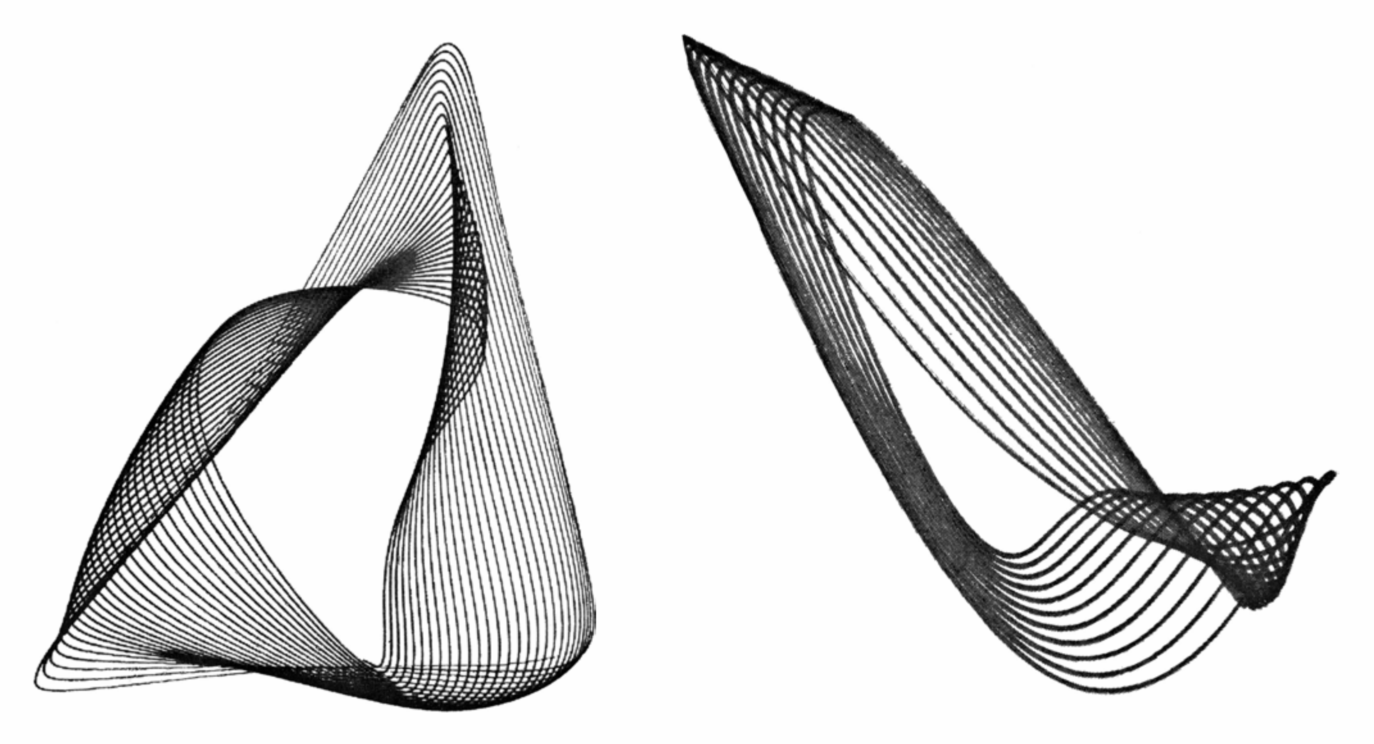
En general, los armonígrafos mecánicos usan como osciladores o bien péndulos o bien manivelas. El movimiento de una manivela horizontal, cuando se mira lateralmente, tiene el mismo aspecto que el de un péndulo visto desde abajo. Ambos son ejemplos de movimiento armónico. Las máquinas a base de péndulos son más fáciles de construir y tienen la ventaja adicional de producir figuras que van disminuyendo en tamaño a medida que la amplitud va decreciendo.

Una versión simple de máquina de péndulo consiste en una linterna suspendida de una cuerda o de un delgado alambre. El bulbo de la lámpara está tapado y la luz sólo puede salir a través de un orificio del tamaño de la punta de un alfiler. Si se hace oscilar la lámpara de modo elíptico y se fotografía desde abajo con un cierto tiempo de exposición se obtiene como imagen sobre la placa una elipse espiralada. El eje de la elipse gira a causa de las inevitables irregularidades de la suspensión.

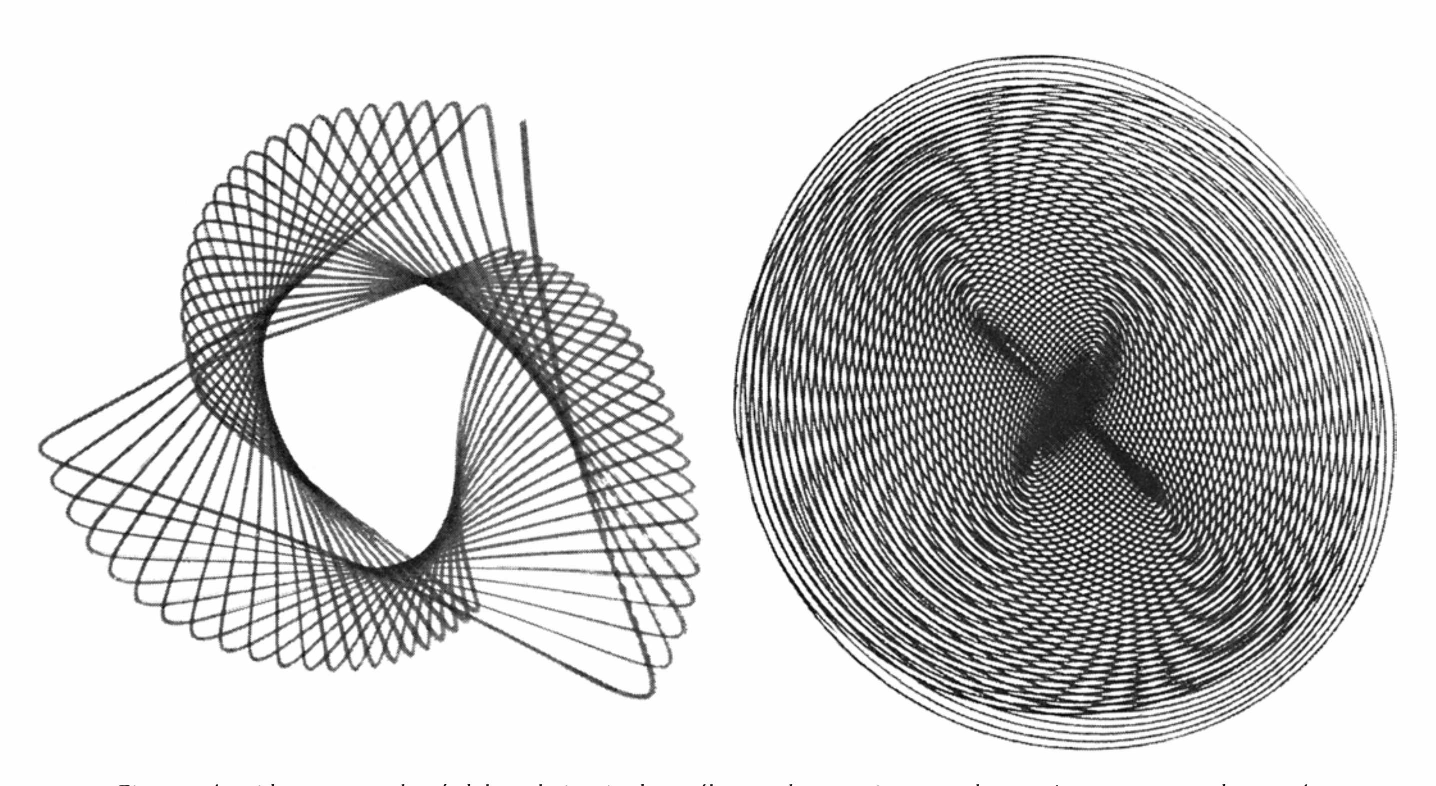
Un péndulo algo más complicado, que genera figuras semejantes a las de Lissajous, se debe al físico británico Hugh Blackburn. Consiste en un peso que cuelga de dos delgados alambres de igual longitud sujetos a puntos algo separados de la barra de suspensión. Mediante una anilla, que puede deslizarse, los alambres se mantienen juntos en un punto convenientemente elegido. De este modo, los dos alambres forman en parte un cable único cuya longitud puede variarse sin más que correr el anillo. En los museos científicos se exhiben con frecuencia péndulos de este tipo que tienen como pesa un embudo de orificio pequeño por el que cae arena fina que deja un registro del movimien-



Un armonígrafo de manivela



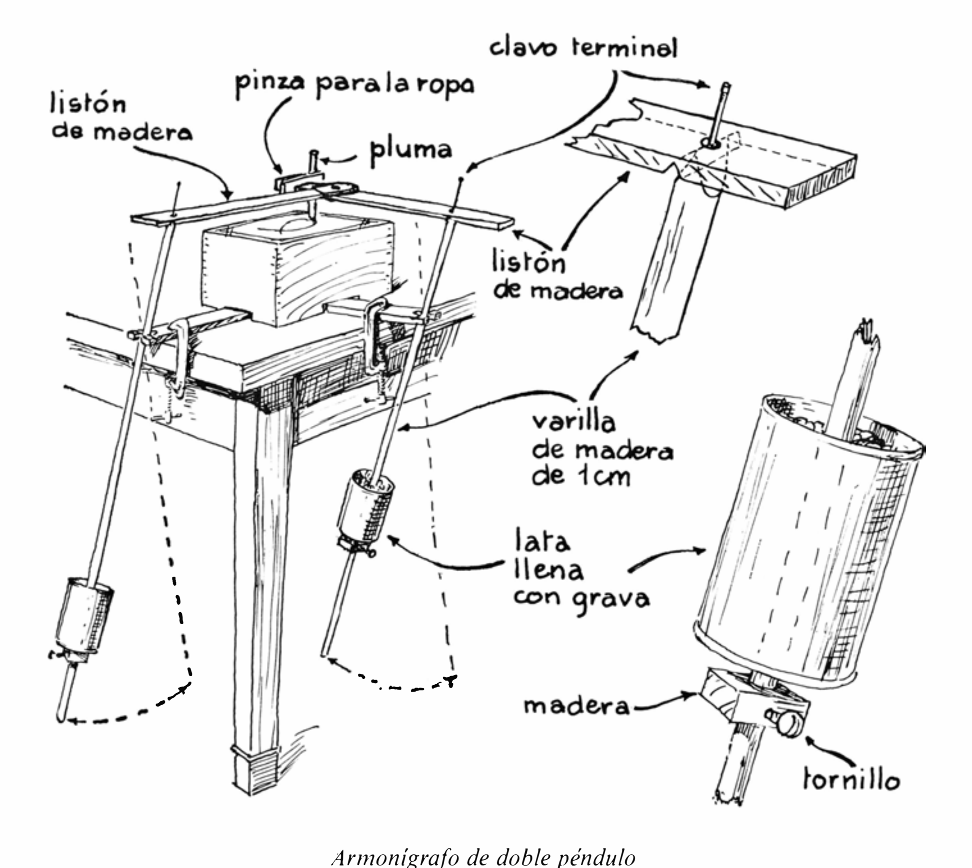
Figuras obtenidas con un armonigrafo de doble péndulo elíptico



Figuras obtenidas con un solo péndulo: a la izquierda, análogas a las anteriores y, a la derecha, con aspecto de moaré

to, en la superficie horizontal situada debajo.

Con madera pueden construirse fácilmente máquinas simples del tipo de manivela. Una de ellas, construida por Tom Barnard de Tucson, Arizona, tiene como base una tabla de madera de 45 × 60 cm de unos 2 cm de grueso. La tabla sostiene tres poleas conectadas por una sola correa *(véase la ilustración de la pág. 106)*. La cara superior de la polea mayor, que tiene unos 25 cm de diámetro, sirve de apoyo a una hoja de papel que gira sujeta a ella. Un brazo ajustable lleva la pluma y es accionado por otra polea más pequeña. La máquina proporciona sus dibujos más interesantes cuando el diámetro de esta polea no es submúltiplo del diámetro de la polea mayor. El extremo de este brazo va conectado a uno de los agujeros dispuestos con radios variables en la polea tractora. Tal disposición permite al operador regular a voluntad la acción del brazo. Su otro extremo cabalga sobre una de las muescas existentes en el borde del otro lado de la tabla soporte; la muesca empleada determina el radio del armonigrama. Un bolígrafo corriente está sujeto a la varilla mediante un puente metálico y



Armonígrafo de doble péndulo

Aparato con doble péndulo elíptico

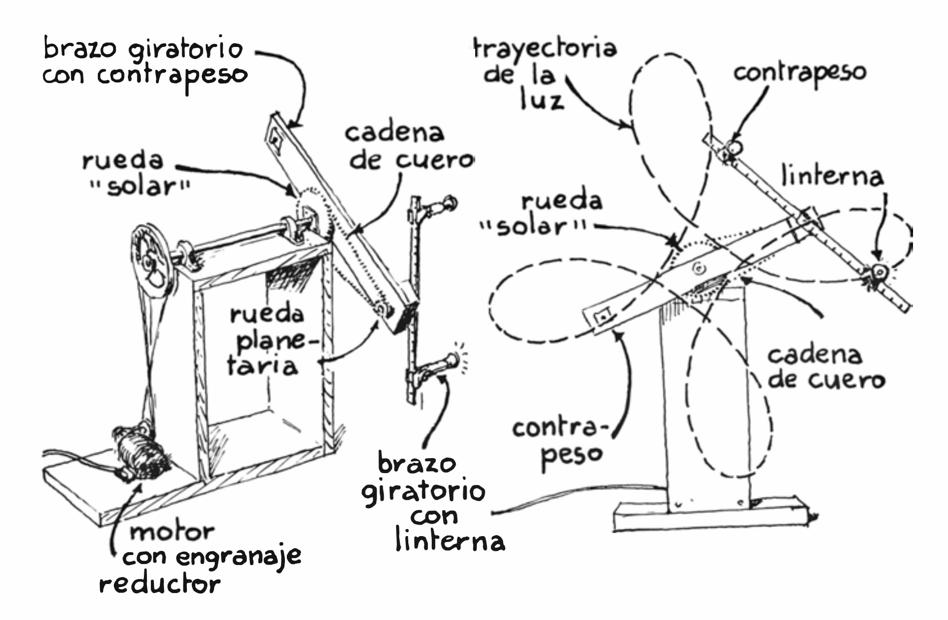
una gomita. El dispositivo es accionado a mano mediante una tercera polea.

Una máquina algo más complicada, pero menos portátil, es el armonígrafo de péndulos gemelos *(véase la ilustración superior izq.)*. Estos, mediante una articulación, van unidos a la pluma de modo independiente y oscilan en planos perpendiculares entre sí. Una pequeña mesa sostiene el papel sobre el que se mueve la pluma. El período de cada péndulo puede ajustarse subiendo o bajando la pesa correspondiente. Hemos fabricado nuestros péndulos con varillas de madera de algo más de 1 cm de diámetro y de cerca de un metro de longitud. Las varillas pivotan sobre cuchillas situadas unos 25 cm por debajo del extremo superior. Las pesas de los péndulos consisten en simples latas llenas de grava que pesan unos 3 kilogramos.

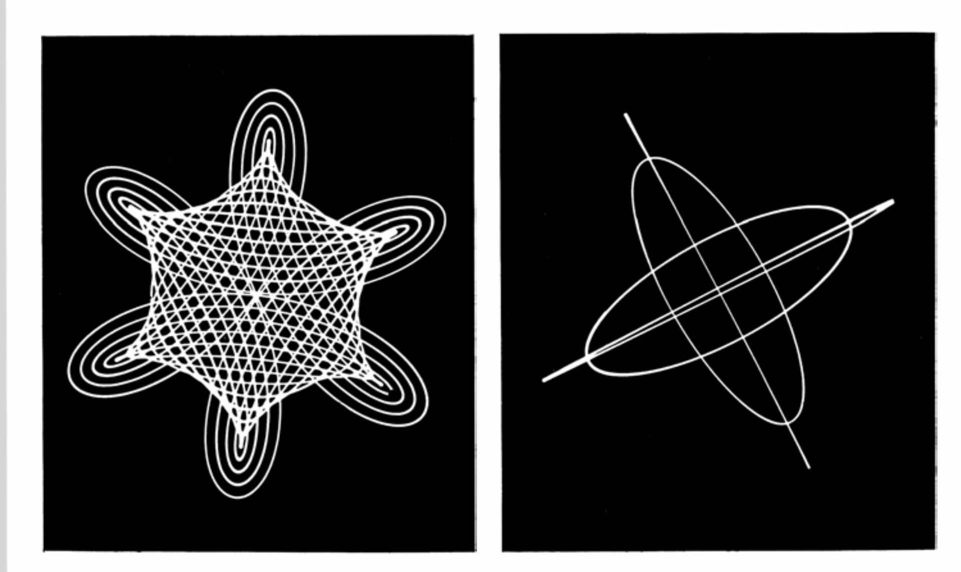
Hemos observado que el bolígrafo escribe mejor si se lastra con 50 g de arandelas de hierro. El dispositivo proporciona figuras con líneas bastante espaciadas pues el bolígrafo frena rápidamente al conjunto. Las curvas resultan más juntas si el bolígrafo se sustituye por una pluma tubular de dibujo (Leroy). La pluma ejerce una presión de sólo unos pocos gramos. Se aumenta su capacidad de tinta colocando en su parte superior un tubito de algo más de un centímetro de longitud y de unos 5 mm de diámetro. Empleamos las plumillas de los tamaños 0 y 1.

Nuestra máquina más moderna es una versión, muy elegante por su simplicidad, del péndulo elíptico doble descrito por primera vez en 1907 por experimentadores británicos. Consiste en una pequeña plataforma de madera lastrada con unos 22 kg de hierro *(véase la ilustración inferior izq.)*. La plataforma está suspendida del techo mediante un yugo y un largo tubo de hierro con una cardan o cojinetes de bolas. Las curvas se dibujan sobre una hoja de papel, fijo sobre la plataforma. La pluma está unida a uno de los extremos de un brazo de palanca equilibrado. El dispositivo descrito sólo genera circunferencias, elipses y líneas rectas. El interés reside en que pueden superponerse dibujos idénticos, algo corridos uno respecto del otro, de modo que aparecen figuras con aspecto de moaré.

Las posibilidades de la máquina pueden ser potenciadas del modo más sorprendente suspendiendo por debajo de la plataforma un segundo péndulo menos pesado mediante una nueva cardan. Cuando el péndulo inferior oscila con arreglo a su propio período natural per-



Un armonígrafo productor de hipocicloides



Figuras obtenidas con un armonígrafo generador de hipocicloides

turba el movimiento de la plataforma oscilante de un modo muy variado. Las curvas más sugestivas se producen cuando la frecuencia del péndulo superior está en una relación sencilla con la del inferior, tal como 3:2 o bien 2:1. Los factores ajustables que determinan una figura son: las dos amplitudes del péndulo superior, el desfase entre ellas, el desfase entre los dos péndulos, la fase y amplitudes del péndulo inferior y la razón de frecuencias entre los dos péndulos. Apenas hemos empezado a estudiar la gran variedad de figuras que pueden ser producidas con la máquina del doble péndulo elíptico.

Son igualmente interesantes las máquinas que funcionan a base de manivelas, especialmente las que generan hipocicloides y epicicloides. Una hipocicloide es la curva descrita por un punto de una circunferencia que rueda por la parte interior de otra circunferencia; en la epicicloide, la rodadura se realiza por la parte exterior de la otra circunferencia. Las máquinas de este tipo pueden ser ajustadas para generar toda una familia de curvas que incluye la cardiode, el caracol de Pascal, la deltoide, la astroide, el trébol, el trébol de cuatro hojas, la elipse, la circunferencia e incluso la línea recta.

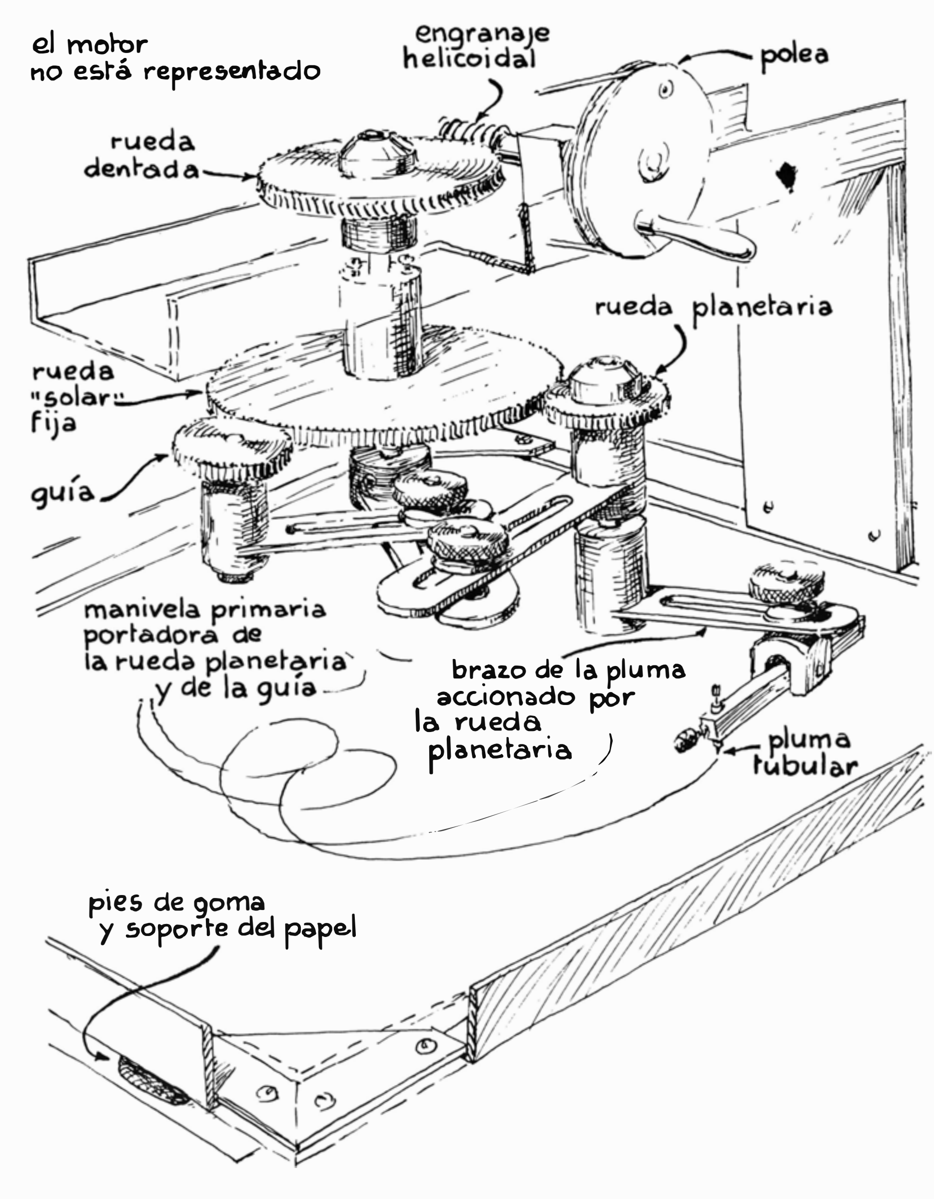
Una máquina sencilla a base de manivela ha sido construida por Everett Clement de Peterborough, New Hampshire. E. Clement realizó el trabajo para estudiar las curvas que aparecen en la pantalla de un osciloscopio electrónico cuando

se conectan a sus terminales dos oscilaciones eléctricas de forma y fase adecuadas. Su dispositivo consiste en un brazo giratorio accionado por un motor eléctrico. En un extremo lleva un segundo brazo que gira a su vez mediante una rueda dentada planetaria y una cadena que engrana con una rueda dentada del eje fijo ("Sol").

En un extremo de este segundo brazo se encuentra una pequeña linterna, cuya luz es fotografiada con una exposición larga en una habitación oscura. El primer brazo describe una circunferencia, mientras la linterna, que marca un punto de la segunda circunferencia, genera hipocicloides. El radio de la figura puede ajustarse alterando la posición de la linterna en su brazo soporte.

Una máquina más compleja y de mayores posibilidades ha sido fabricada por H. A. Cata, de Ginebra, Suiza. Su autor le ha dado el nombre de ciclógrafo *(véase la ilustración de la pág. 110)*. Puede generar tanto hipocicloides como epicicloides. Además de las figuras que pueden obtenerse con la máquina de Clement, el ciclógrafo demuestra que la cardiode, el caracol de Pascal y la nefroide son simplemente casos especiales de la epicicloide. Sin embargo, las espirales, las cúbicas, las cuárticas y demás curvas de orden superior caen fuera de sus posibilidades actuales.

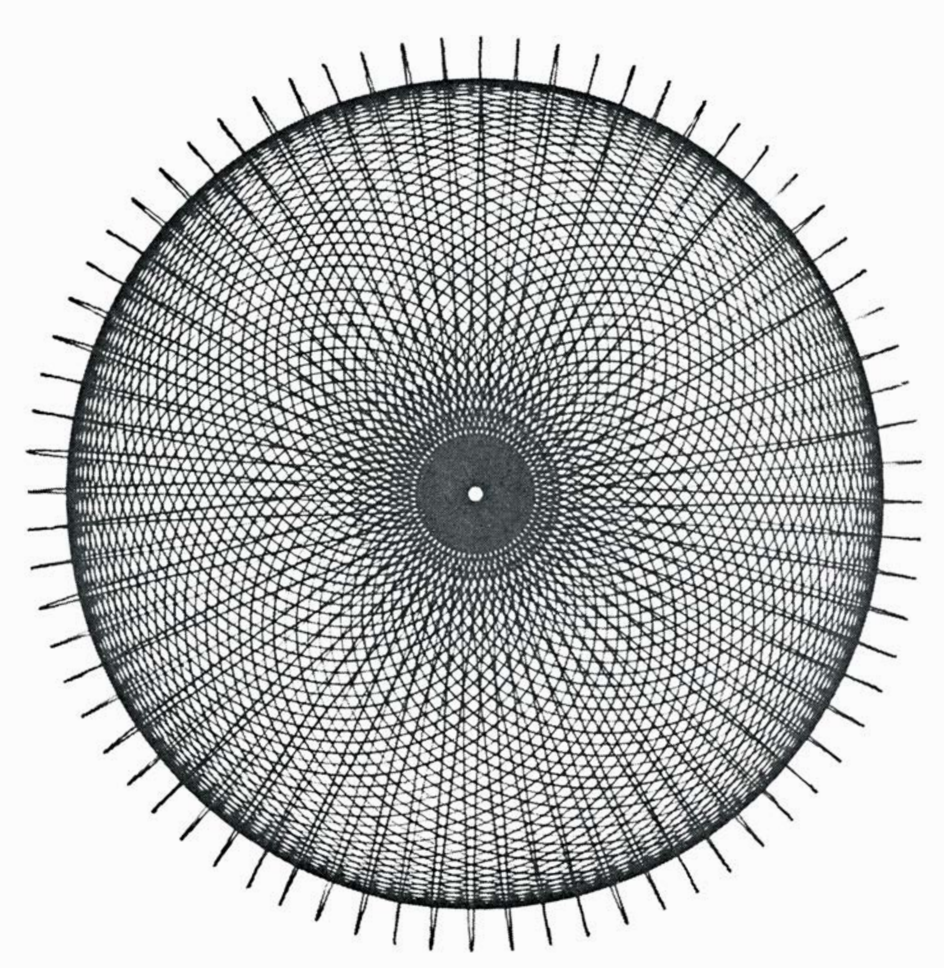
A excepción de las máquinas con un sólo péndulo, las demás pueden emplearse para producir armonigramas estereoscópicos, es decir, curvas que parecen tridimensionales al observarlas mediante un estereoscopio o con la ayuda de un prisma. Con alguna práctica, hay quienes son capaces de ver las figuras en tres dimensiones sin más que cruzar los ojos de tal modo que el ojo derecho vea sólo la figura izquierda y el izquierdo la derecha. Para obtener parejas es-



Ciclógrafo.

tereoscópicas con una máquina de doble péndulo *(ilustración superior de la pág. 108)*, se empieza por registrar una figura que se obtiene soltando a la vez los dos péndulos. La segunda figura, que formará pareja con la anterior, se obtiene soltando un péndulo algo después que el otro. De este modo, el comienzo de la curva de uno diferirá en unos cinco grados respecto a la del otro. Las dos figuras resultantes han de montarse luego, una junto a la otra, con una distancia entre centros de unos 5 cm, para lo que se reducen de tamaño por un procedimiento fotográfico. Cuando se montan, su orientación debe ser tal que coincidieran si fuesen corridas y superpuestas. Los armonigramas estereoscópicos pueden conseguirse también con máquinas a base de engranajes, simplemente aumentando el radio del brazo que lleva la pluma en un 5 por ciento cuando se genera la segunda figura.

Hemos presentado varias versiones de armonígrafos fáciles de construir y hemos dado ejemplos de las figuras que son capaces de proporcionar. Sin embargo, el tema no quedaría completo si no mencionáramos los principios físicos y matemáticos fundamentales que están en la base de tales diseños. El movimiento armónico simple tiene una interpretación bien sencilla: es el resultado de proyectar sobre un diámetro las posiciones de un punto que recorre una circunferencia con velocidad constante. La posición del punto proyectado está determinada en cualquier instante por el radio de la circunferencia multiplicado por el seno trigonométrico del ángulo descrito por el punto proyectante. Este ángulo es igual a la velocidad angular multiplicada por el tiempo transcurrido. Este tipo de movimiento es rápido cerca del centro de la circunferencia y se va haciendo gradualmente más lento hasta pararse por un instante en el extremo. Acto seguido vuelve a acelerarse hacia el centro, para después reducirse de nuevo a cero en el extremo opuesto del diámetro. Tal movimiento debió serle ya familiar al hombre de las cavernas cuando se colgaba de un árbol: como lo es, hoy en día, a los chicos que se columpian. Pueden citarse otros ejemplos como el movimiento de una pesa que cuelga de una cinta de goma, o diversas vibraciones, tales como las de una cuerda de violín, las compresiones de aire en los instrumentos de viento a base de tubos y las de las partículas subatómicas responsables de la emisión de ondas electromagnéticas. En resumen, el sencillo movimiento tan maravillosamente representado en los armonigramas es un elemento básico de todos los fenómenos naturales.



Hipocicloides superpuestas a epicicloides

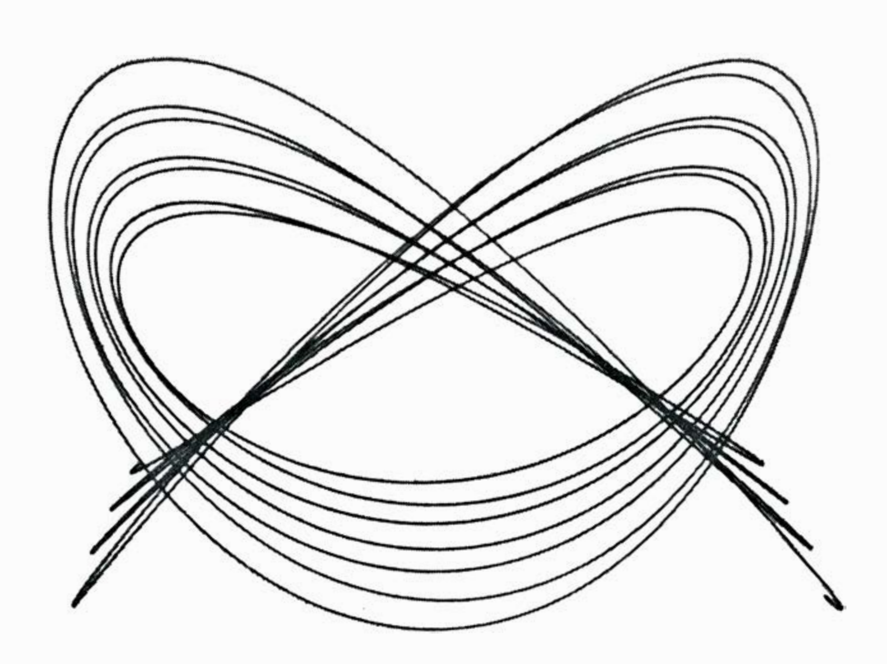


Figura obtenida con un péndulo de Blackburn

Libros

Un espejo de la evolución que ha sufrido la lingüística generativa

Violeta Demonte

ESTUDIOS DE GRAMATICA GENERATIVA. Reunidos por Víctor Sánchez de Zavala. Barcelona, Labor, 1976. Hablando laxamente podemos decir que esta obra es, en lo que a metodología respecta, un espejo de la evolución que ha sufrido la lingüística generativa desde su aparición hasta los tiempos más recientes.

Es así que encontramos aquí artículos típicos de lo que podríamos denominar su *primera generación,* esto es, que tratan en detalle de la construcción de reglas de estructura sintagmática y transformatorias, buscan su motivación e intentan analizar la capacidad predictiva de las mismas. Otros, acordes con actitudes lógicamente posteriores, estudian propuestas alternativas en relación con ciertos fenómenos y tratan de escoger entre ellas en función de razones metateóricas (su simplicidad, rigor, o redundancia, o el poder de los dispositivos descriptivos que utilizanw consecuencias empíricas (a dónde conduce el análisis en cuestión y qu4 otros fenómenos puede o no caracterizar) o capacidad explicativa (hasta qué punto esos análisis sirven para poner en relación principios generales y dan razón, así, en cierto modo, de la naturaleza de la mente humana). Alguno, por fin, investiga las propiedades semánticas y pragmánticas de determinadas construcciones y postula tipos de representación formal en los que podrían integrarse esas propiedades.

La consideración, por otra parte, de las áreas o temas gramaticales que se abordan en esta antología muestra nítidamente hasta qué punto es cierto aquello de que los hechos dependen de las teorías. Aunque algunos de los estudios se refieren a temas clásicos de las gramáticas de las lenguas que abordan –ejemplos de ello serían los relativos al pronombre catalán "en" y a los determinadores del castellano– es evidente que su reanálisis con los instrumentos descriptivos (más la exigencia de rigor metodológico) que proporciona la nueva lingüística arroja luces y pistas insospechadas en torno a la naturaleza y razón de ser de estos hechos lingüísticos. Pero, además, se introducen aquí asuntos que son por completo novedosos o que apenas tienen tradición en nuestros estudios gramaticales como el tema de la presuposición en relación con las oraciones completivas, la vinculación entre ciertas construcciones incoativas y causativas o el por qué y el carácter de las restricciones que afectan a las secuencias de clíticos del castellano. Es verdad también que en el tratamiento de todas estas cuestiones se observa la dependencia –harto explicable sin lugar a dudas– en que aún nos encontramos con respecto a la actividad lingüística que tiene lugar en Estados Unidos. En efecto, con alguna excepción excepcional (como puede ser el artículo del compilador) los puntos de vista desde los que se suscita el análisis de los hechos de nuestras lenguas han sido todos adoptados previamente en relación con fenómenos del inglés, lo cual, como es bien sabido, en nada invalida (puesto que la teoría pretende caracterizar lo que es común a todas las lenguas humanas) que se los tome en cuenta para estudiar fenómenos de otras lenguas, aunque sí sea un indicio de que nos falta el ambiente y la tradición necesarios para descubrir nuevos aspectos tanto de la teoría que utilizamos como de las lenguas sobre las cuales trabajamos.

Esas perspectivas a las que acabo de aludir son las que tendré en cuenta para presentar y discutir los artículos de esta antología. A riesgo de ser imprecisa pero con el fin de no caer en una casuística abrumadora los he agrupado en cinco secciones que, a mi juicio, responden a las cuestiones centrales que se debaten en los mismos (aunque quizá dejen de lado aspectos interesantes de ellos). En la presentación de los trabajos procuro situarlos en el debate teórico dentro del cual emergen, y que les da sentido; en la discusión me centro fundamentalmente en una crítica interna (con criterios internos a nuestra disciplina), sobre todo metodológica. En todo caso, procuro explorar las consecuencias de estos trabajos y de destacar, dentro de la brevedad de una recensión, lo que de programático pueda haber en ellos.

El orden de los apartados que siguen pretende reproducir ese orden cronológico abstracto de aparición de conjeturas, réplicas y contrarréplicas en la lingüística generativo-transformatoria.

1) Un enfoque transformatorio de la pronominalización

En la primera etapa de la lingüística que nos ocupa, en la cual se pretendía que la gramática podía tener una *base* relativamente simple y concreta y que el peso de la explicación debía confiarse a las transformaciones sintácticas, se trataron con tales dispositivos fenómenos que luego, al hilo de la polémica *generativismo/interpretativismo,* fueron considerados como semánticos (y por tanto como no propios del subcomponente transformatorio) por los propios chomskianos ortodoxos. Un caso típico es el de los pronombres, respecto de los cuales se mantienen, hoy por hoy, una hipótesis transformativista y otra interpretativista. El hecho es, de todos modos, que ambas soluciones parecen empíricamente equivalentes.

Joan Argente [JA], "Un exercici d'anàlisi transformacional entorn del pronom EN", realiza un estudio de este pronombre partitivo, adnominal y adverbial del catalán, y de los hechos relacionados con el proceso de pronominalización. Su trabajo consta de dos partes. La primera (# 2-4) es una descripción y análisis de los tipos de oraciones en que puede aparecer "en"; en esta sección, y simultáneamente con dicha descripción, se van sugiriendo las reglas que podrían dar razón de los fenómenos lingüísticos que se estudian. En la segunda (# 5) se formalizan las reglas antes introducidas [a saber, DISLOCACIÓ, PRO-en, DAVANT-en, DELISIÓ DEL TEMA (además de otras de alcance mayor como Permutació de subjecte indeterminant)] y se determina su orden de aplicación. En esta segunda parte se establece también la configuración de la estructura subyacente a la cual deben aplicarse las reglas mencionadas. El trabajo concluye con dos breves apartados acerca de las implicaciones teóricas de su análisis.

La característica quizá más destacada

de este trabajo es que es excesivamente abarcador, en el sentido de que alude en 33 páginas a 8 reglas transformatorias. (Es verdad también que ello se justifica plenamente dado el objetivo pedagógico que se asigna). Aunque las construcciones que estudia son bastante restringidas, el autor alude frecuentemente a procesos relacionados con el de formación del 'en', los cuales quedan, en algunos casos, escasamente planteados o incluso apenas justificados. Cito como ejemplo la interpolación que va de la p. 21 a la 23 acerca de la sintaxis de orden frente a la de relación, o el texto de # 4.5 acerca de la posibilidad o imposibilidad de que 'en' pueda representar a un SP dependiente del sujeto, en el que la manera más especulativa que argumentativa de presentar los problemas hace que tal cuestión no quede claramente especificada. (Hago notar, dicho sea de paso, que en Ruwet y Kayne el análisis del fenómeno equivalente en francés lleva a la conclusión de que este 'en' tiene una derivación transformacional distinta de la del 'en' proveniente de la posición posverbal.)

El artículo de JA es rico en ideas originales y, por otra parte, las reglas más pertinentes que propone (DISLOCACIO y PRO-en) parecen del todo plausibles. Lo que parece faltarle a este trabajo son aseveraciones independientes demostrativas, esto es, argumentos, sobre las cuales pueda fundarse certeramente la asignación de esas reglas a esos hechos lingüísticos. Más concretamente, los hechos se describen en los mismos términos en que luego se van a formular las reglas y no se mencionan otras propiedades del material que prueben la necesidad de derivarlo de ese modo. Es verdad que en la segunda parte se busca justificación independiente para algunas transformaciones, y se demuestra muy acertadamente la presencia de 'de' en la estructura subyacente, y con ello se subsana algo esta deficiencia, pero las dos transformaciones esenciales no se motivan suficientemente, aunque, quizá formen parte de la gramática del catalán y estén implicadas en estos procesos.

II) Los clíticos: ¿condiciones eductivas o análisis transformacional?

No es exagerado afirmar quizá que el tema de los clíticos es el que más trabajos ha suscitado en la literatura transformativista relativa al castellano. La cuestión, efectivamente, es crucial y permite elegir entre concepciones alternativas, bien definidas, de la teoría y de la estructura lingüística: ¿es conveniente restringir el funcionamiento de las reglas de la gramática (por medio de principios sobre la aplicación de las mismas o de filtros positivos y estáticos que eliminen las secuencias mal formadas derivadas a través del componente sintáctico) sin cambiar su forma, o resulta más adecuado constreñir aquéllas y no incorporar nuevos dispositivos? La adopción de la primera alternativa da lugar a gramáticas más poderosas y, en consecuencia, a teorías menos estrictas acerca de la naturaleza subyacente de los fenómenos lingüísticos; la segunda, al menos en apariencia, define mejor la noción de gramática posible y, por tanto, de lenguaje humano posible. Como siempre, la elección entre ambas alternativas es, en gran medida, una cuestión empírica.

Perlmutter ve como inevitable la introducción de *restricciones sobre la estructura superficial* o, lo que es lo mismo, de *condiciones eductivas sobre las transformaciones.* Salvador Bastida [SB] en su "Restricciones de orden en las secuencias de clíticos del castellano: dos condiciones" argumenta de manera brillante, articulada y convincente a favor de una solución transformativista. El razonamiento de SB se basa en la comprobación de que el filtro de Perlmutter engloba condiciones heterogéneas, no sólo de orden sino también semántico-cuantitativas y relativas a las transformaciones particulares que afectan a ciertos clíticos. Sus argumentos se refieren, en primer lugar, a la insuficiencia (puesto que debe completarse con una especificación de *casos* si no quiere dejar pasar secuencias agramaticales como '*Yo te le acerco') y redundancia (puesto que al exigir que las secuencias sean estrictamente monótonas y crecientes repite restricciones semántico-cuantitativas que deben especificarse en otras partes de la gramática) del filtro de Perlmutter. En segundo lugar, muestra que es falso, que SE deba encabezar toda secuencia de clíticos. Estas comprobaciones conducirían no sólo a reformular el cedazo en cuestión sino también a modificar, y a restringir a ciertos casos, la aplicación de la regla que crea el falso SE, de la cual depende crucialmente el filtro de Perlmutter. Así las cosas, resulta más lícito entender a éste como un conjunto de transformaciones de reordenación. SB concluye su trabajo con una discusión del orden que debe establecerse entre las transformaciones relativas al SE y la ordenadora de toda secuencia de clíticos.

El razonamiento es riguroso e inobjetable desde el punto de vista lógico. Sólo quiero, pues, discutir algunas de las complicaciones que, a mi juicio, se siguen de la solución formal que escoge y señalar a la vez las implicaciones teóricas del enfoque adoptado. La regla ordenadora que propone SB (*vid.* 46, p. 95) hace uso de nociones mixtas: las nociones funcionales de DT y AC y la semántica, o semántico-sintáctica, de CR (correferencial), ninguna de las cuales, si nos atenemos a la *teoría normal ampliada* (Chomsky), pertenece al componente sintáctico. Las reglas interpretativas del componente semántico son las que, según Jackendoff, dan razón de las relaciones *temáticas* (en la estructura profunda) y de las de correferencia (al final de cada ciclo sintáctico). Por tanto, si aceptamos la segunda teoría de Chomsky, y si la regla de SB es la única posible, esa información temática (o funcional) y correferencial debería ser suministrada al componente sintáctico, debería quedar patente en él, antes de que opere la regla (46). (Es razonable argüir, sin embargo, que las nociones de caso podrían representarse por medio de *rasgos* léxicos adscritos a las formas pronominales. Hay que recordar, sin embargo, que ese recurso, aparte de *ad hoc,* es redundante puesto que en la base ya se distingue, por posición y construcción, el acusativo del dativo.)

Por otra parte, la regla (46) tiene la desventaja de ser excesivamente amplia (como el autor dice es, en realidad, varias reglas, con lo cual nos alejamos del objetivo de que las transformaciones deban efectuar operaciones simples y muy bien delimitadas (Chomsky, Emonds), dentro de un enfoque *preservador de la estructura,* propone también una solución transformativista para los clíticos del francés. En ese análisis se postulan tres reglas de colocación de los clíticos a través de las cuales éstos quedan situados en la configuración que sería exigida por la restricción eductiva. Una mayor investigación acerca de aquellas implicaciones y de estas posibles soluciones permitiría, quizá, perfeccionar las ya seguramente definitivas aportaciones del análisis de SB.

III) Sintaxis abstracta: construcciones incoativas y causativas

Las primeras críticas a la *teoría normal* se orientaron a considerar el estatuto que la semántica ha de tener en la descripción lingüística. La concepción que surge de estas posturas críticas, la

semántica generativa, propondrá que la representación subyacente de las oraciones sea semántico-sintáctica, en concreto, que los árboles sintácticos puedan terminar en elementos de naturaleza tanto sintáctica como semántica. Las consecuencias de esta posición son varias e importantes, al menos a primera vista. Una de ellas es la incorporación a la estructura subyacente de formantes abstractos (en especial verbos) que no tienen reflejo fonológico alguno en la estructura superficial, otra la desaparición o fusión de las categorías sintácticas clásicas, una tercera el aumento de complejidad de la descripci6n sintáctica que, por ejemplo, puede representar como compuesta a una oración simple de la estructura superficial. Estas tres caracterSsticas definen lo que se ha denominado *sintaxis abstracta*, cuya abstracción es, en última instancia, una función del número de transformaciones que venga a requerir.

Vinculada a estos planteamientos surge la hipótesis de la *descomposición léxica*, esto es, la idea de que en la estructura subyacente debe representarse el significado, descompuesto, de las piezas léxicas. Desde el punto de vista del tratamiento de los elementos del léxico esta hipótesis no propone nada que vaya más allá del análisis componencial clásico, aunque sí haya servido para investigar cuáles pueden ser algunos de los elementos semánticos *primitivos* que intervienen en el significado de las palabras. Desde el punto de vista de la teoría su importancia proviene de constituir un argumento a favor de la no existencia de un nivel aislado de estructura profunda, a tenor de la existencia de transformaciones previas a la inserción del léxico.

III. 1 Josep Nadal [JN]: "Sintaxi abstracta i derivació: transitivitat *versus* intransitivitat en el català actual", realiza una investigación acerca de la estructura profunda de ciertas oraciones copulativas, incoativas y causativas, y de las relaciones entre ellas. Propone JN que las oraciones incoativas del tipo de

(1) *a.* La muntanya es fa verda
b. La muntanya verdeja

tienen una estructura profunda intransitiva en la cual el sujeto es una oración insertada: 'La muntanya (és) verda". A las estructuras subyacentes correspondientes a (1*a*) y (1*b*) (que sólo difieren en el carácter, respectivamente, −ó +PRO del verbo matriz) se aplican una serie de transformaciones independientemente motivadas a través de las cuales se obtienen las estructuras superficiales mencionadas. Las oraciones causativas similares a

(2) *a.* En Joan va fer que l'aigua bullir
b. En Joan va fer bullir l'aigua

se originan en una estructura profunda transitiva con un sujeto agente y un objeto directo oracional igual en estructura a la correspondiente a (1). Para resolver el problema que plantean las piezas léxicas irregulares recurre a la convención del *marcado* del léxico introducida por Lakoff.

El análisis es prolijo, es, sin duda, descriptivamente adecuado, y aplica al catalán lo propuesto por Lakoff (1965) para el inglés y por Roldán (1971) para el castellano.

Aparte de la réplica que, en términos generales, puede hacerse a este análisis por el recurso a un procedimiento más descriptivo que explicativo como es el marcado de las piezas léxicas, y por la necesidad inescapable en que se ve de apelar a transformaciones *ad hoc* (las transformaciones INCOATIVA Y CAUSATIVA serían un ejemplo) para así eliminar los elementos abstractos que introduce, querría referirme a ciertos datos, predicciones y asignaciones de estructura que, desde mi punto de vista, deberían ser tenidos en cuenta por quien quisiese elegir entre esta solución o alguna otra para proceder a explicar las construcciones de este tipo. (Se me disculpará por presentar ejemplos del castellano, pero creo que en casi todos los casos pueden encontrarse equivalentes en catalán.)

Obsérvese, en primer lugar, que si se adopta la hipótesis de la O insertada (*vid.* fig. (8), p. 158) la representación subyacente que se obtiene caracterizará del mismo modo a estas oraciones que a las que efectivamente llevan subordinadas subjetivas. Pero lo que hace más problemática esta asignación de estructura es que esa incrustación no parece tener justificación sintáctica independiente. Prueba de ello, como ha señalado Nápoli para el caso del italiano, es que a esa O sujeto no puede aplicarse transformación cíclica alguna aparte de la que asciende su sujeto (la cual es necesaria, por lo que respecta a estructuras como éstas, sólo si se adopta la hipótesis en cuestión).

En segundo lugar, este análisis predice que los causativos transitivos tienen un correspondiente incoativo [ya que la estructura subyacente incoativa aparece como objeto directo del 'fer" causativo (*vid.* fig. [42a], p. 170)]. En castellano al menos, existen muchos contraejemplos para esa correlación: 'adornar la casa" (hacer que la casa esté adornada)/ *la casa se adorna. (*Cf.* Luján para otra problemática predicción relativa a la transformación de RELATIVIZACION).

Por otra parte, la gran asistematicidad que encontramos cuando analizamos las correlaciones entre estas construcciones (incoativos a los que no corresponden causativos: 'atragantarse'/* 'atragantar' y viceversa (*vid.* más arriba), la no equivalencia semántica estricta entre ciertos incoativos y transitivos léxicamente idénticos: 'incapacitarse'/'incapacitar', el hecho de que existan piezas semánticamente incoativas: 'aparecer ', 'fallecer", etc., que ni cumplen las características sintácticas propias de los incoativos reflexivos, ni tienen construcciones correspondientes copulativas y causativas, son tres elementos de juicio que, desde mi punto de vista, deben tenerse en cuenta para completar el análisis, en principio no insatisfactorio, de estas construcciones.

III. 2. El artículo de Ignacio Bosque [IB], "Sobre la interpretación causativa de los verbos adjetivales", retoma la hipótesis de la descomposición léxica. Este trabajo difiere de los del resto de la antología en tener un objetivo más descriptivo que probatorio y en referirse del léxico. El autor estudia, entre otras cuestiones que quedan más sugeridas que analizadas, la estructura semántica de verbos relacionados con algunos de los adjetivos que admiten grados de comparación (en concreto, los adjetivos polares): 'encarecer'/'abaratar', 'alargar'/ 'acortar', etc., e insinúa que el adverbio 'más' debería aparecer en esa estructura preléxica vinculado, dentro de cada par, a los adjetivos que 'admiten un grado intensivo cuando la cualidad primitiva sufre un cambio en cierto modo ascendente (negativo a positivo) o como quiera que lo llamemos (*op. cit.* página 109).

El material lingüístico que se presenta es interesante y está bien seleccionado. Creo, no obstante, que si se retoma el tema de la causatividad en castellano habrá que tener en cuenta las críticas que ha recibido el tratamiento sintáctico de este problema, independientemente de que en el pueda tener repercusiones importantes en el lexicón. En efecto, se han aducido argumentos semánticos (la diferencia de valor veritativo que hay entre las oraciones con CAUSAR + V y aqué-

llas en las que aparece un verbo simple (Chomsky; Ruwet); semántico-sintácticos (relativos a las restricciones selectivas y de coaparición; (Ruwet, Hall Partee), y sintácticos [acerca de la aplicabilidad de ciertas transformaciones (Fodor, Shibatani)] –aparte las alusiones al peligro de que esa estructura preléxica pueda ser crecientemente compleja (Chomsky), punto sobre el cual el trabajo que consideramos vendría a aportar nuevos datos que arrojan serias dudas acerca no sólo de la adecuación sino de la aceptabilidad del análisis causativo.

IV La naturaleza de las presuposiciones y su representación en la gramática

Un tema de amplia tradición en lógica y filosofía, el de las presuposiciones, ha pasado a ser, en los últimos años, motivo central de las teorizaciones de los lingüistas. Dos trabajos de la antología que comentamos abordan el tema de la presuposición. El de Conxita Lleó [CL]: "La presuposición y los verbos factivos en castellano" y el de Angel Manteca [AM]: "Nota sobre la presuposición en castellano". En el primero se trata el problema de la representación de la presuposición en la descripción lingüística, en el segundo la relación que guarda con los modos del verbo.

IV. 1. Paul y Carol Kiparsky estudiaron un tipo de presuposición que consiste en que el hablante da por supuesta la verdad de la completiva incrustada a ciertos verbos que denominaron *factivos* (ejemplos de ellos son: 'darse cuenta de', 'lamentar', etc). Las oraciones con estos verbos llevan asociadas ciertas peculiaridades sintácticas que no se dan cuando el verbo de la matriz es no-factivo. Para explicar estas diferencias sintácticas los Kiparsky formularon la hipótesis de que la presuposición de la completiva debe reflejarse en la estructura generada por la *base.* CL pone en duda la adecuación y suficiencia de esta forma de representación, crítica que había iniciado Rivero y que ha continuado, por ejemplo, Kempson.

La objeción que aporta este trabajo se centra en la pérdida de factividad (el no conllevar presuposiciones) de algunos de los verbos factivos cuando aparecen insertos de *cierta* manera en el complemento de "verbos creadores de mundo' (p. ej. 'soñar', 'imaginar' etc). CL encuentra que los verbos creadores de mundo no siempre dejan pasar la presuposición a la oración matriz y que esa ausencia de presuposición depende de que el complemento del verbo factivo haya aparecido previamente, y en la misma oración compleja, como aseveración dependiente de un verbo creador de mundo [*vid.* ejs. (32) y (33)]. La observación parece atinada. La solución que propone CL distingue en la estructura subyacente de estas oraciones el contenido aseverado del contenido presupuesto. La autora aplica así a las completivas castellanas una idea que se venía desarrollando, en relación con otros temas, en trabajos de la misma época (Cf. Hooper, y Hooper y Terrel): la de distinguir formalmente entre la aserción y la presuposición.

IV. 2. AM en su breve nota sobre la presuposición afirma que la diferencia de estructura profunda, en términos de presuposición o no presuposición, defendida por Rivero para explicar la aparición de indicativo (presuposición de que el complemento es verdadero) o subjuntivo (ninguna presuposición) es innecesaria porque los verbos que admiten cualquiera de los dos modos ('suponer', 'admitir', 'confiar' (ejs. de AM) son ' realizativos' [performatives] y por tanto actúan como ' tapones' [plugs] (Karttunen), bloqueando todas las presuposiciones de la oración complemento. Creo que en la nota de AM se dan pruebas muy débiles para lo que se quiere demostrar, fundamentalmente porque se confunden cuestiones conceptuales. Mi afirmación se basa en la siguienes razones.

En primer lugar, los verbos que AM menciona son de un carácter realizativo muy dudoso. 'Admitir', 'suponer' o 'sospechar' no es realizar una *acción verbal* puesto que estos verbos no llevan ninguna fuerza inlocutiva, en el sentido de Austin. Compáreselos, por ej., con 'jurar', 'prometer' o 'declarar'. Obsérvese también que no satisfacen ninguno de los requisitos sintácticos que se asocian a los realizativos (*vid.* Vendler), no son, pongamos por caso, verbos que cuando usados en presente de indicativo describan una acción puntual, antes bien se refieren a un lapso indefinido de tiempo. Los verbos que admiten cualesquiera de los modos son, casi siempre, verbos de *actitud proposicional* (*vid.* Vendler y Karttunen), también llamados de *opinión* (Terrel). Hasta aquí, mi réplica parece puramente terminológica. El problema que debemos plantearnos es si estos verbos cancelan o no la presuposición de la subordinada. Karttunen encuentra difícil decidir al respecto y, a través de varios ejemplos, muestra que en realidad pueden actuar de las dos maneras. (En el mismo sentido, la oración (20b) de AM es, para mí, presuposicional).

Pero aun cuando aceptásemos que estos verbos no llevan (o bloquean) presuposiciones (*vid.* Hooper y Terrel que los clasifican como *asertivos,* esto es, no presuposicionales) tal decisión sería irrelevante para poner en tela de juicio el análisis de Rivero. Como ha sido ya señalado (*vid.* Klein, Rivero), aunque no lo explicite, emplea el término presuposición en un sentido distinto del estándar, tanto si pensamos en una presuposición de índole semántica como en las pragmáticas (si es que esta distinción tiene sentido). Así, en los casos en que Rivero dice que la completiva está presupuesta: [Los corredores creen [que el belga ganó la carrera]$_B$]$_A$, no puede referirse a que A implique lógicamente que B ni a que la negación de A: 'Los corredores no creen que el belga ganó la carrera" también implique que 'El belga ganó la carrera" es verdadera. (Más aún, las oraciones en las que, según Rivero, el hablante presupone la verdad del complemento son casi siempre no-presuposicionales, en sentido normal, y las que no presuponen corresponden a predicados *factivos* y *emotivos,* esto es, a predicados que llevan presuposiciones.

Tampoco es verdad que la diferencia entre una construcción con indicativo y la misma con subjuntivo sea *siempre* de "mayor o menor incertidumbre del hablante". Considérese el par: 'Entiendo que te disgusta su llegada'/'Entiendo que te disguste su llegada'; ¿hay aquí una diferencia de *grado* de incertidumbre o dos juicios de tipo diferente (comprobación frente a participación)?

Las consideraciones teóricas del apartado # 3 son tan excesivamente generales e incongruentes que resultan difíciles de comentar.

V) Propiedades semánticas y parámetros pragmáticos

El centramiento en la sintaxis, por muy positivo que sea o haya podido ser para el descubrimiento y caracterización de los mecanismos formales y universales que subyacen a la actividad lingüística, es indudable que acotó de manera drástica el tipo de cuestiones que parecían intuitivamente abordables por los programas de investigación lingüística. El movimiento semantista surgió de la conciencia de esas limitaciones y aunque no haya conseguido quebrar el paradigma generativista volvió a poner

sobre el tapete temas clásicos del estudio del lenguaje tales como el del contexto en que tiene lugar la comunicación verbal, el problema de la referencia o el del conocimiento previo o presupuesto que aporta el hablante, los cuales, de un lado, y como era de esperar, han desembocado en otros ya no tan reducibles a aquel paradigma y, de otro, han servido para dar lugar a modos de reflexionar sobre los hechos algo diferentes. Se afirma que la gramática debe reflejar la intuición del hablante, pero ¿la intuición acerca de que? La sintaxis transformacionalista nos ha acostumbrado a analizar nuestra intuición acerca de las realizaciones –semánticas y sintácticas– en que se encuentran las oraciones, pero nos ha alejado de la posibilidad, empleada ya antes por los filósofos, de analizar, preteorética o formalmente, cómo expresamos por medio del lenguaje nuestra relación con el mundo. (Entendiendo por mundo, claro es, algo bastante limitado como puede ser la presencia de un interlocutor, el conocimiento acceso de que disponemos hacia los objetos que mentamos, lo que queremos obtener, diciendo o implicitando, con nuestro acto verbal.)

V. 1. Víctor Sánchez de Zabala [VSZ] es, que yo sepa, el único lingüista español que se haya adentrado por estos nuevos y prometedores caminos. En "Sobre una ausencia del castellano", incluida en esta antología que compila, va a averiguar a qué condiciones semánticas y pragmáticas obedece el fenómeno superficial de ausencia del *determinador.* El trabajo tiene toda la densidad, originalidad y penetración que caracteriza a la producción de este investigador y constituye, sin duda, una contribución importante no sólo para el esclarecimiento de la naturaleza y funcionamiento de los SSNN sino para la caracterización de la actividad que despliega el sujeto parlante.

Dicho brevemente, VSZ discute las propiedades relativas al modo de hacer referencia, de presentar información y apelar al oyente de los SSNN sin determinador, propiedades que pueden descubrirse gracias a la introducción de los conceptos pragmáticos de *referencia mentadora,* de *accesibilidad para el oyente* y de *selectividad.* La aparición de las construcciones sin determinador con dichas propiedades se alinea con ciertas características sintácticas de los SSNN así como de los VV. El trabajo termina con una propuesta, y ejemplos (que, posiblemente, no se presentan con toda la información previa que hubiera hecho falta para el lector no muy familiarizado con este sector de la teoría), relativa al formalismo adecuado para representar tanto el contenido proposicional como el conocimiento del hablante.

El trabajo, por lo que tiene de programático y de sugestivo más que de demostración acabada, ha de producir, es indudable, nuevos resultados en el estudio del castellano. Me atrevo a sugerir, por ejemplo (y esto es, desde todo punto de vista, obvio) que habría que investigar en qué otras regiones del lenguaje y con qué restricciones o peculiaridades actúan esos parámetros pragmáticos, y qué nuevas propiedades pueden derivarse de ellos. Parece pertinente, por ejemplo, y esto está ya iniciado en el trabajo de VSZ, enlazarlos con el estudio sobre el empleo de los modos indicativo y subjuntivo, y de las relativas restrictivas y no-restrictivas.

V. 2. En R. Lakoff se presentan algunos hechos lingüísticos en torno a la coordinación y se sugiere, no se pretende demostrar, que en toda coordinación simétrica debe funcionar una factor semánticopragmático: el *tema común,* el cual, o bien estará presente en forma patente en la estructura superficial o podrá derivarse a partir de una combinación de presuposiciones (de diversos tipos) y deducciones realizadas por el hablante sobre la base de los dos elementos antes mencionados. Juan Luis Tato [JLT], en "Sobre la coordinación", revisa las sugerencias de Lakoff, las critica y propone que la condición debe sustituirse por la de *co-temporalidad* –indicada una por la propia R. Lakoff (*vid. op. cit.* pp. 117 y 125)– esto es, que la aceptabilidad de la coordinación depende de que los verbos tengan un esquema temporal-aspectual homogéneo. Esta condición, no obstante, parece que tampoco es suficiente, a juzgar por oraciones como su (53), y debería complementarse con una consideración del carácter (concreto o abstracto) de la acción verbal. Que este criterio adicional también es insuficiente se ve tras oraciones como 'María labra su porvenir y su hermano vende palomitas en la esquina", perfectamente gramatical en determinados contextos.

A juzgar por este trabajo parece que la noción de *tema común,* y la de *compatibilidad semántica,* tienen que perfilarse todavía algo más y que su configuración definitiva va a depender de cuáles sean exactamente las condiciones que aquéllas incluyan y qué jerarquía pueda establecerse entre éstas.

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

LA ERRADICACION DE LA VIRUELA

SMALLPOX. C. W. Dixon. J. & A. Churchill Ltd, 1962.

SMALLPOX. Donald A. Henderson en *Maxcy-Rosenau Preventive Medicine and Public Health: Tenth Edition,* dirigido por Philip E. Sartwell. Appleton-Century-Crofts, 1973.

SMALLPOX ERADICATION IN WEST AND CENTRAL AFRICA. William H. Foege, J. D. Millar y D. A. Henderson en *Bulletin of the World Health Organization,* vol. 52, n.º 2, págs. 209-222; 1975.

GENERACION FOTOVOLTAICA DE ELECTRICIDAD

DIRECT USE OF THE SUN'S ENERGY. Farrington Daniels. Yale University Press. 1964.

SATELLITE SOLAR POWER STATION. P. E. Glaser in *Solar Energy,* vol. 12, n.º 3, págs. 353-361; 1969.

GROWTH OF CONTROLLED PROFILE CRYSTALS FROM THE MELT, PART III: THEORY. Bruce Chalmers, H. E. LaBelle, Jr., y A. I. Mlavsky en *Materials Research Bulletin,* vol. 6, n.º 8, págs. 681-690; agosto, 1971.

ESTUDIOS SOBRE LOS RIBOSOMAS MEDIANTE DISPERSION DE NEUTRONES

NEUTRON SCATTERING. B. P. Schoenborn y A. C. Nunes en *Annual Review of Biophysics and Bioengineering,* vol. 1, págs. 529-552; 1972.

NEUTRON SMALL-ANGLE SCATTERING OF BIOLOGICAL MACROMOLECULES IN SOLUTION. Heinrich B. Stuhrmann en *Journal of Applied Crystallography,* vol. 7, part. 2, págs. 173-178; 1 abril 1974.

DETERMINATION OF QUATERNARY STRUCTURE BY SMALL ANGLE NEUTRON SCATTERING. D. M. Engelman y P. B. Moore en *Annual Review of Biophysics and Bioengineering,* vol. 4, pág. 219-241; 1975.

NEUTRON SCATTERING MEASUREMENTS OF SEPARATION AND SHAPE OF PROTEINS IN 30S RIBOSOMAL SUBUNIT OF *ESCHERICHIA COLI:* S2–S5, S5–S8, S3–S7. Donald M. Engelman, Peter B. Moore y Benno P. Schoenborn en *Proceedings of the National Academy of Sciences of the United States of America,* vol. 72, n.º 10, págs. 3888-3892; octubre, 1975.

ERUPCIONES DE RAYOS GAMMA COSMICOS

OBSERVATIONS OF GAMMA-RAY BURSTS OF COSMIC ORIGIN. Ray W. Klebesadel, Ian B. Strong y Roy A. Olson en *Astrophysical Journal Letters,* vol. 182, n.º 2, part. 2, págs. L-85-L88; 1 junio 1973.

RECENT OBSERVATIONS OF COSMIC γ-RAY BURSTS. T. L. Cline en *Annals of the New York Academy of Sciences: Seventh Texas Symposium on Relativistic Astrophysics,* vol. 262, págs. 159-163; 15 octubre, 1973.

A PRELIMINARY CATALOGUE OF TRANSIENT COSMIC GAMMA-RAY SOURCES OBSERVED BY THE *VELA* Satellites. Ian B. Strong, Ray W. Klebesadel y Roy A. Olson en *Astrophysical Journal Letters,* vol. 188; n.º 1, part. 2; págs. L1-L3; 15 febrero, 1974.

OBSERVATIONS OF γ-RAY BURSTS. I. B. Strong. W. D. Evans and R. W. Klebesadel en *Annals of the New York Academy of Sciences: Seventh Texas Symposium on Relativistic Astrophysics,* vol. 262, págs. 145-158; 15 octubre, 1975.

HOLOGRAMAS DE LUZ BLANCA

PHOTOGRAPHY BY LASER. Emmett N. Leith y Juris Upatnieks en *Scientific American,* vol. 212, n.º 6, págs. 24-35; junio, 1965.

ADVANCES IN HOLOGRAPHY. Keith S. Pennington en *Scientific American,* vol. 218, n.º 2, págs. 40-48; febrero, 1968.

LASERS AND LIGHT: READINGS FROM SCIENTIFIC AMERICAN. W. H. Freeman and Company, 1969.

HOLOGRAPHY: WITH AN INTRODUCTION TO THE OPTICS OF DIFFRACTION, INTERFERENCE, AND PHASE DIFFERENCES. H. Arthur Klein. J. B. Lippincott Company, 1970.

ORGANIZACION SOCIAL DE LOS MACACOS JAPONESES

PRIMATE SOCIETIES: GROUP TECHNIQUES OF ECOLOGICAL ADAPTATION. Hans Kummer. Aldine Publishing Company, 1971.

THE SOCIAL BEHAVIOR OF MONKEYS. Thelma Rowell. Penguin Books, 1972.

CONCEPTUAL AND METHODOLOGICAL PROBLEMS ASSOCIATED WITH THE STUDY OF AGGRESSIVE BEHAVIOR IN PRIMATES UNDER SEMINATURAL CONDITIONS. G. Gray Eaton en *The Neuropsychology of Aggression,* dirigido por Richard E. Whalen. Plenum Press, 1974.

TEMPESTADES DE POLVO

A COMPARISON BETWEEN LABORATORY AND ATMOSPHERIC DENSITY CURRENTS. J. E. Simpson en *Quarterly Journal of the Royal Meteorological Society,* vol. 95, nº. 406, págs. 758-765; octubre, 1969.

EFFECTS OF THE LOWER BOUNDARY ON THE HEAD OF A GRAVITY CURRENT. John E. Simpson en *Journal of Fluid Mechanics,* vol. 53, part. 4, págs. 759-768; 27 junio 1972.

AN AMERICAN HABOOB. S. B. Idso, R. S. Ingram y J. M. Pritchard en *Bulletin of the American Meteorological Society,* vol. 53, nº. 10, págs. 930-935; octubre, 1972.

LA QUIMICA PERCLOROORGANICA

CHLOROCARBURES AROMATIQUES ET ALCHYLAROMATIQUES. SYNTHÈSES ET PROPRIÈTÉS, M. Ballester, en *Bulletin de la Société Chimique de France,* n.º 1; págs. 7-15; 1966.

INERT CARBON FREE RADICALS, M. Ballester, en *Pure and Applied Chemistry,* vol. 15, págs. 123-151; 1967.

POLYCHLOROAROMATIC COMPOUNDS, M. Ballester y S. Olivella, Editado por H. Suschitzky, Plenum Press; 1974.

JUEGOS MATEMATICOS

AMUSEMENTS IN MATHEMATICS. Henry E. Dudeney. Dover Publications, Inc., 1958.

THE BEST MATHEMATICAL PUZZLES OF SAM LOYD. Dirigido por Martin Gardner. Dover Publications, Inc., 1959.

Indice anual

Este índice contiene los nombres de los autores y artículos aparecidos en Investigación y Ciencia *a partir de octubre de 1976. Se incorpora asimismo la relación alfabética de las secciones* Juegos matemáticos y Taller y laboratorio.

AUTORES

ARTICULOS

JUEGOS MATEMÁTICOS

TALLER Y LABORATORIO